

Academic Vocabulary Unleashed potential?

A corpus study of English course materials for
advanced Norwegian learners of English



Kimberly Skjelde

Master's Thesis

Department of Foreign Languages

University of Bergen

November 2015

Acknowledgements

I would like to give a special thanks to my supervisor Torill Hestetraet for her extensive insight, untiring willingness to help during this entire process, kind words of encouragement and her ability to make me believe in what I was doing. I would also like to thank my fellow MA students for many interesting discussions and their heartfelt encouragement. Thanks especially to Jaspreet Gloppen and Christin Beenfelt for being my critical friends.

I would not have dared take on the challenge of this endeavor without the love and support of my family. Thanks to my mother for always encouraging her daughters to learn more about the world around them. Thanks also goes to my children who have patiently put up with a mother that never seems to stop “taking classes”. I owe the biggest thanks of all to my loving husband who is always willing to lend an ear, and never fails to encourage me when I need it the most; your support has been invaluable.

Finally, to all of my English students, I thank you for inspiring me to want to learn more and hope that you will someday experience this same kind of inspiration.

Kimberly Skjelde
Foldnes
November 2015

Abstract in Norwegian

Innlæring av ord og begreper utgjør en vesentlig del av språkinnlæring, også for fremmedspråk. Som engelsklærer har jeg ofte vært frustrert når jeg føler at elever ikke klarer å bruke formelle ord i passende situasjoner. Målet med dette studiet har vært å undersøke den formelle språkbruken i faktatekster funnet i engelske læreverker for vg1 studiespesialiserende elever. Det har også vært et mål å undersøke om disse elevene vil kunne tilegne seg akademiske ord ved å lese faktatekster på egenhånd. Jeg har definert generelt akademisk ordforråd ved bruk av en ofte brukt internasjonal ordliste, the Academic Word List (AWL).

En analyse av hvordan et akademisk ordforråd er brukt i faktatekster publisert i læreverker og hvilken innvirkning disse kan ha på innlæring av engelsk ordforråd er sentral i studiet. Derfor har de teoretiske rammene vært tuftet på «usage-based» teorier som forklarer denne innlæringsprosessen ut fra det perspektivet at mennesker lærer språk gjennom å bruke det. De teoriene som jeg fremhever er relatert til betydningen av skriftlige kilder som input, betydningen av repetisjon i innlæringsfasen og betydningen av at innlæringen først skjer når språklige former blir lagt merke til.

Opgaven er en korpusstudie av 21 faktatekster tatt fra tre ulike læreverker. Syv tekster fra hver bok er analysert. Tekstene omhandler to tema relatert til kompetansemålene i engelskfaget, nemlig engelsk som et globalt språk og urbefolkningsgrupper.

Resultatene viser at flertallet av de akademiske ordene i liten grad er brukt i tekstene og ikke vil kunne fremme læring bare ved at elever leser tekstene alene. Funnene viser også at elever i liten grad er eksponert for generelt akademisk språk i tekstene skrevet for læreverkene, noe som tyder på at det er viktig å kombinere bruken av disse med skriftlige autentiske tekster. Samtidig viser resultatene at autentiske tekster ikke nødvendigvis er vanskeligere å forstå, selv om de bruker et mer akademisk språk. Studiet støtter nyere forskning som viser at generelt akademisk språk som definert gjennom AWL, i stor grad består av ordfamilier som er høyfrekvente og som elevene derfor ofte vil møte i autentiske diskurser. Funnene viser også at akademisk ordforråd i liten grad er brukt i glosser. Dette styrker grunnlaget for å si at det kan være nødvendig å bruke en liten del av undervisningen, også på videregående nivå, til undervisning relatert til innlæring av akademiske ord.

Table of Contents

Acknowledgements	iii
Abstract in Norwegian	iv
Table of Contents	v
List of Abbreviations.....	viii
List of Tables.....	ix
List of Figures	x
List of Appendices	xi
1. Introduction	1
1.1 Aim and Scope	1
1.2 Why General Academic Vocabulary?	3
1.3 Why Reading Factual, Textbook Texts?	6
1.3.1 The Importance of Textbook Texts.....	6
1.3.2 Reading and L2 Vocabulary Acquisition	7
1.4 Research Questions	10
1.5 Outline of the Thesis	10
2 Theoretical Background	12
2.1 General Academic Vocabulary	12
2.1.1 The Academic Word List (AWL)	13
2.2 Corpus Linguistics.....	16
2.2.1 Corpora.....	17
2.2.2 Counting Words	18
2.2.3 Using CL to Determine Frequency Levels.....	19
2.3 Usage-based Theory	20
2.4 Relevant Hypotheses	22
2.4.1 The Role of Input	22
2.4.2 The Frequency Hypothesis.....	26
2.4.3 The Role of Noticing.....	30
2.4.4 The Lexical Quality Hypothesis.....	32
3. Methods and Materials	37
3.1 Materials.....	37
3.2 Methods.....	37
3.3 Research Design	39
3.4 Choice of Materials	40
3.4.1 Textbook Choice	40

3.4.2 Text Choice	41
3.5 Text Analysis.....	42
3.5.1 Token, Type and Word Family	42
3.5.2 Text Preparation	43
3.5.3 VocabProfiler (VP) Classic.....	45
3.5.4 VocabProfiler (VP) Compleat.....	46
3.5.5 Combining Both Programs.....	47
3.5.6 Range Analyses	47
3.5.7 Glossary Analyses	48
3.6 Ethical Issues.....	50
3.7 Reliability and Validity	50
3.8 Limitations	53
4. Results	53
4.1 AWL Vocabulary Use.....	53
4.1.1 Percentage of Total Text	54
4.1.2 In-text Frequency	58
4.1.3 Range.....	62
4.1.4 AWL One Occurrence.....	64
4.2 Glossing.....	67
4.2.1 Total Glossary Coverage	68
4.2.2 AWL Glossary Coverage	69
4.3 Lexical Coverage.....	74
4.3.1 General Lexical Coverage	74
4.3.2 AWL Lexical Coverage	79
4.3.3 In-depth Investigation of One Text	82
5. Discussion of Results	84
5.1 Brief Overview	84
5.2 AWL Vocabulary Use.....	85
5.2.1 AWL Text Coverage	85
5.2.2 Range Frequency	87
5.3 Glossing.....	90
5.3.1 Glossary Coverage	90
5.4 Lexical Coverage.....	92
5.4.1 General Lexical Coverage	92
5.4.2 AWL Lexical Coverage	94

5.5 Brief Summary of Findings	96
6. Conclusion.....	98
6.1 Key Findings	99
6.1.1 AWL Usage.....	99
6.1.2 Glossing.....	99
6.1.3 Lexical Coverage.....	100
6.2 Contributions	101
6.3 Implications	101
6.3.1 Materials design	101
6.3.2 Classroom practices.....	102
6.4 Recommendations for Further Study	102
References	104

List of Abbreviations

AWL	The Academic Word List
BNC	British National Corpus
CL	Corpus Linguistics
COCA	Corpus of Contemporary American English
EAP	English for Academic Purposes
EFL	English as a Foreign Language
GSL	General Service List
IELTS	International English Language Testing System
NSD	Norwegian Social Science Data Services
L2	Second Language
SLA	Second Language Acquisition
VP	VocabProfiler

List of Tables

Table 1. Total corpus average of AWL used per text for tailored texts	54
Table 2. Total corpus average of AWL used per text for authentic texts	55
Table 3. Access to English: Total average of AWL per tailored text	55
Table 4. Access to English: Total average of AWL per authentic texts	56
Table 5. Stunt: Total average of AWL per tailored text	56
Table 6. Stunt: Total average of AWL per authentic texts	57
Table 7. Targets: Total average of AWL per tailored texts	57
Table 8. Targets: Total average of AWL per authentic texts	57
Table 9. AWL word family in-text frequency of six or more repetitions.....	58
Table 10. Access to English: List of AWL headwords and word types, six plus.....	60
Table 11. Stunt: List of AWL headwords and word types, six plus	60
Table 12. Targets: List of AWL headwords and word types, six plus	61
Table 13. AWL word families occurring across topic related texts, total corpus	62
Table 14. Percentage of in-text AWL word families used once	65
Table 15. Percentage of AWL word families used once across three and four texts	66
Table 16. AWL word families: BNC/ COCA frequency levels in each textbook	66
Table 17. AWL word families: BNC and COCA frequency levels across topics	66
Table 18. Average glossary coverage for the corpus	68
Table 19. Glossary coverage for tailored texts in Access to English	68
Table 20. Glossary coverage for tailored texts in Stunt	69
Table 21. Glossary coverage for tailored texts in Targets	69
Table 22. Total AWL glossary coverage per textbook	70
Table 23. Per textbook: AWL glossary coverage with one occurrence	72
Table 24. BNC and COCA frequency levels for entire corpus	75
Table 25. Access to English: BNC/COCA frequency levels of AWL word families	80
Table 26. Stunt: BNC/COCA frequency levels of AWL word families	81
Table 27. Targets: BNC/COCA frequency levels of AWL word families	81

List of Figures

Figure 1. An Expanded Model of the Lexical Quality Hypothesis	32
Figure 2. Mixed Methods Diagram	40
Figure 3. 98% Lexical Coverage	77
Figure 4. 95% Lexical Coverage	79

List of Appendices

Appendix 7.1.1 Information to the schools	109
Appendix 7.1.2 Overview of replies from schools	110
Appendix 7.2 Text Analyses entire text	111
Appendix 7.2.1. Access to English	111
Appendix 7.2.2 Stunt	203
Appendix 7.2.3 Targets	304
Appendix 7.3 Text analyses – glossary items	413
Appendix 7.3.1. Access to English	413
Appendix 7.3.2 Stunt	421
Appendix 7.3.3 Targets	430
Appendix 7.4 Range analyses	443
Appendix 7.4.1 Access to English	443
Appendix 7.4.2 Stunt	449
Appendix 7.4.3 Targets	458
Appendix 7.5 AWL in-text once	466
Appendix 7.5.1 Textbook: Access to English	466
Appendix 7.5.2 Textbook: Stunt	467
Appendix 7.5.3 Textbook: Targets	468
Appendix 7.5.4 Text only file	469
Appendix 7.6 VP-Compleat analysis of all AWL vocabulary	476
Appendix 7.7 Frequency levels of total text	490
Appendix 7.8 Norwegian Social Science Data Services (NSD)	492

1. Introduction

“Knowledge of things and knowledge of the words for them grow together. If you do not know the words, you can hardly know the thing.”

Henry Hazlitt, *Thinking as a Science*

1.1 Aim and Scope

Vocabulary is a fundamental part of all language acquisition, and no less so in the acquisition of a second language. Since the 1990's the field of second language acquisition (SLA), has placed more focus on the study of vocabulary acquisition. Questions often investigated concern what words second language (L2) learners should learn and how this is done most efficiently. The current study will investigate general academic vocabulary and how it may best be taught.

The reasons for this line of enquiry rest in my personal experience as a teacher for advanced L2 English learners. I have often times felt that my practices in the classroom fall short when it comes to teaching my students to comprehend and use a wider vocabulary. For this reason, I was interested in finding out more about the process of L2 vocabulary acquisition. My decision to focus on L2 vocabulary acquisition through reading textbook texts came from my dependence as a teacher on factual, textbook texts when introducing new, curriculum based topics. During written and oral discussion related to these topics I have often experienced my students struggle to comprehend and use formal language. I therefore wanted to examine how to best facilitate the acquisition of this type of vocabulary.

The aim of this master's thesis is to investigate how general academic vocabulary is used in factual, textbook texts in order to then assess if this is done in such a manner that L2 vocabulary acquisition of general academic vocabulary may be expected. I have, therefore, analyzed seven texts related to two different topics, from three different textbooks. These texts make up the corpus of written input used as the basis for the research analyses. The textbooks are written for use in the last obligatory English course for high school students in

Norway. I will characterize the student target group as advanced L2 learners. English is taught as a second language in Norway and it would be as appropriate to use the term English as a Foreign Language (EFL) learners, but I have chosen to use the more commonly used term of L2 learners. There is a debate about the use of English in Norway being so prominent the English can be considered a second language, but I will not debate the matter further. I have termed the students as advanced English learners even though this is a term often related to university level students. I have done so because these students have had obligatory English lessons for all ten years of their education before starting high school.

When asking questions related to what vocabulary to teach and how, linguists largely agree that L2 learners are best served by learning the most commonly used vocabulary items first (Nation, 2013). Researchers have looked to authentic texts in order to comprise lists over the most frequently used words in English. The General Service List (GSL) was developed by West (1953), and is still in use today. However, with the development of computer programs aiding the process, many new lists have been developed. One of the latest developments is Nation's BNC-COCA frequency list made up of 29 word family groups (2012). A word family is "a headword, its inflected forms, and ...closely related derived forms" (Nation, 2013, p. 11). Nation's list is based on word families present in both the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). This list can be used to determine how frequently a word family is used in English i.e., the frequency levels of the words in a text. The BNC-COCA corpora is made up of millions of spoken and written English words used in current authentic situations (Nation, 2012). These lists, among others, have been used to set vocabulary goals for L2 learners.

The GSL, a list over the most frequent 2,000 English word families has long been seen as an appropriate vocabulary learning goal for beginning L2 learners. The next logical vocabulary goal for L2 learners seeking higher education has been to learn academic vocabulary (Coxhead, 2006; Gulden, 2008; Nation, 2013). In general terms, academic vocabulary refers to vocabulary prevalent in texts used for academic purposes (Baumann & Graves, 2010; A. Coxhead, 2006; Nation, 2013; N. Schmitt, 2010). While this progression is still seen as appropriate, new research has suggested that the GSL + academic vocabulary is not enough for students to comprehend completely what they are reading (Cobb, 2010; Nation, 2006; D. Schmitt & Schmitt, 2012). For the current study, general academic vocabulary will be investigated more closely. Here, general academic vocabulary has been operationalized with a word list created by Coxhead in 2000, called the Academic Word List

(AWL). A more recent frequency list created by Nation in 2013, the BNC-COCA frequency lists, will also be used to examine in more detail the frequency of AWL vocabulary used in the corpus for my study.

Corpus based studies such as this are dependent upon corpora and frequency lists to conduct data analyses. Research for the current study will be conducted with the use of the VocabProfiler and Range programs (Cobb, 2015) and manual analyses. The study will be conducted using instruments producing quantitative data, but will also be heavily supplemented with a qualitative approach towards population selection, some data collection and in relation to the discussion of results.

The scope of this study is rather broad because I have chosen both to analyze the use of academic vocabulary in written texts and do so in such a manner that can help determine if this usage may promote vocabulary acquisition.

1.2 Why General Academic Vocabulary?

Internationally, vocabulary researchers often make a distinction between words that are academic and those that are technical, referring to academic vocabulary as “common in different kinds of academic texts” as opposed to technical vocabulary “consist[ing] of words that are closely related to the content of a particular discipline” (Nation, 2013, pp. 19,303). Some researchers make this same distinction with the use of the terms general versus discipline-specific academic vocabulary (Heibert & Lubliner, 2008; Nagy & Townsend, 2012). Other researchers argue that academic vocabulary must be defined only in relation to each subject it applies to, thus questioning the idea of general academic vocabulary (Hyland, 2011; Hyland & Tse, 2007). I support the assumption that there is a general academic vocabulary for the English language on the grounds of corpus research. I have chosen to use Coxhead’s Academic Word List (AWL) to operationalize the term (see section 2.1.1).

In a Norwegian context, one important English subject competence aim for first year, general studies, high school students is to learn to “express oneself fluently and coherently in a detailed and precise manner suited to the purpose and situation” (Utdanningsdirektoratet, 2013). This ability is often described in textbooks in relation to differences between formal and informal English language use (Areklett, Hals, Lindaas, & Tørnby, 2009; Balsvik, Bratberg, Henry, Kagge, & Pihlstrøm, 2015; Burgess & Sørhus, 2013). An important factor in these discussions includes focus on vocabulary use. Here formal vocabulary is often defined

in terms of concrete vocabulary uses such as avoiding the use of personal pronouns and contractions, as well as using a precise, wide vocabulary, i.e. “often longer words or words with origins in Latin and Greek” (Balsvik et al., 2015, p. 32).

Internationally, formal English is often defined in terms of the use of academic language. In an article written for the British Council, David Park defines formal English as “mainly used in writing...academic in tone and...commonly used in academic textbooks, most university essays, business letters and contracts” (n.d.). Nation describes the AWL as “to a large degree a marker of formal written language” (2013, p. 294).

The acquisition of general academic vocabulary is seen by most researchers as being an invaluable part of any student’s education process (Corson, 1997; Gardner & Davies, 2013; Lesaux, Keiffer, Kelley, & Harris, 2014; Nagy & Townsend, 2012). Many also proclaim the need for teaching advanced L2 English learners general academic vocabulary (Coxhead, 2006; Gardner & Davies, 2013; Nation, 2013; D. Schmitt & Schmitt, 2012; Simpson-Vlach & Ellis, 2010). Nation outlines four main reasons: it is “common to a wide range of academic texts, and not so common to non-academic texts..., accounts for a substantial number of words in academic texts..., is generally not known as well as technical vocabulary..., [and] is the kind of specialised vocabulary that an English teacher can usefully help learners with” (2013, pp. 291-293). Coxhead also points out that proficiency in academic vocabulary will give college students with English as an L2 the opportunity to “be part of the academic community” and will be expected of them if they wish to become successful in their studies (2006, p. 3).

In the Norwegian national curriculum, the use of the term academic vocabulary appeared for the first time in the English translation of the national curriculum revisions for the English subject in 2013. According to these aims, students must learn to “understand and use an extensive general vocabulary and an academic vocabulary related to one’s education programme” (Utdanningsdirektoratet, 2013). In my experience this term is not widely discussed or used in relation to classroom teaching in Norway. This despite the expressed requirement for the school system to “provide a foundation for further education...” (Utdanningsdirektoratet, 2006) and the recognition that there is a need for high school students to learn English because this language is increasingly used in higher education (Utdanningsdirektoratet, 2013).

Researchers in Norway also express the need for academic vocabulary acquisition among students seeking higher education. In a study of freshmen university students and high school seniors, Hasselgren discovered an overuse of what she coined “lexical teddy bears” in their written production. These are lexical choices made by advanced L2 English language learners in Norway that show a clear influence from their L1. Her findings showed, among other things, that wrong word choices often led to errors in style (1994). In a follow-up study done 10 years later, Mahan (2013) , found that 62 % of the vocabulary mistakes her Norwegian participants made were related to errors in style. The students’ writing showed an overuse of general verbs, colloquialisms and informal vocabulary not suited to the task. The students simplified the language “by using well-known [high-frequency] words characterized by colloquial vocabulary rather than more precise or academic terms” (My translation Mahan & Brevik, 2013, p. 38). The findings Hasselgren and Mahan have presented are supported by a 2014 study conducted through the EF English Proficiency Index. In this quantitative research, including on-line testing of 910,000 adult participants world wide, they found that the participants in Northern Europe “have exceptionally good English skills” however, “many students do not develop an adequate level of academic English to pursue tertiary studies in the language” (Index, n.d.). Associate Professor, Ann T. Gulden sees the need to “expand the [English for Academic Purposes] EAP portfolio for more categories of students, since they are graduating into an increasingly internationalized society” (2008, p. 207). She continues by saying that “Norwegians are well schooled in general English, but there are aspects of EAP teaching in which we would do well to cooperate at a national level to improve academic English” (2008, p. 208). It is my hope that the current study, by providing new insight into high school students’ exposure to general academic vocabulary in textbook texts, can help find ways of aiding Norwegian students in their acquisition of formal vocabulary.

The process of helping students comprehend and use a more precise, general academic vocabulary is complex and involves both the students themselves, teachers, materials writers and researchers (N. Schmitt, 2008). As such, the current study is limited in focus to one, small part in this process i.e., the general academic vocabulary advanced L2 Norwegian learners of English are exposed to in written course materials designed specifically for these students.

1.3 Why Reading Factual, Textbook Texts?

Ellis and Shintani (R. Ellis & Shintani, 2014) outline conditions needed to promote L2 acquisition based on SLA theory and research. These include access to large amounts of input that is comprehensible to the L2 learner. The input must be used in such a manner that caters to learning and the learners must pay attention to linguistic forms in the input they have not acquired yet (R. Ellis & Shintani, 2014).

Reading is an important source of vocabulary input for L2 learners; however, reading does not necessarily lead to vocabulary acquisition. In order for vocabulary learning to occur implicitly i.e., unintentionally without awareness, the unknown words in a text must be met enough times during reading, and the learner must be able to accurately infer the meaning of words from the context in which they are read (R. Ellis & Shintani, 2014).

The current study aims at providing greater knowledge related to how general academic vocabulary is used in written course materials provided for advanced Norwegian learners of English as an L2. Further, the study aims at investigating whether implicit acquisition of general academic vocabulary is likely to occur during unassisted reading. Even though this is an area that has been widely researched, there remains a need for further studies exploring the “[vocabulary] coverage and potential for vocabulary learning...in English language course books” (Nation & Webb, 2011, p. 171).

1.3.1 The Importance of Textbook Texts

From my own personal experience, teachers are dependent upon texts related to topics outlined in the nation curriculum and these texts are often found in textbooks. A national report discussing research related to the use of course materials in elementary and middle schools across Norway concludes that textbooks continue to dominate in classroom practices (Juuhl, Hontvedt, & Skjelbred, 2010). Several recent studies conducted for the English classroom in Norwegian grade schools confirms the reliance of teachers in English as a Foreign Language (EFL) classes on the use of textbooks. Charboneau found that 61.8 % of the 370 teachers participating in her study “used a textbook as the basis of English reading instruction” (2012, p. 57). Hestetraet has studied teacher cognition in relation to EFL vocabulary acquisition among seventh grade teachers in Norway. Her study also showed the majority of teachers continue to rely on textbooks. Of the 341 respondents to her questionnaire, 92% reported using a textbook, “often or very frequently” (2012, p. 185).

There are two recent, international studies that have investigated vocabulary use in English language teaching (ELT) course books (Matsuoka & Hirsh, 2010; Ruegg & Brown, 2014). The findings in these studies show that there is wide variation when it comes to the use of vocabulary in the analyzed textbooks. In the study conducted by Matsuoka & Hirsh, they examined the use of general academic vocabulary, defined as AWL word families in all 12 texts represented in one textbook. One important finding in their study showed that, on average, over 40% of the AWL word families used in the textbook texts occurred only once (Matsuoka & Hirsh, 2010, p. 64). In a quantitative corpus study conducted in Japan of internationally acclaimed English as a Foreign Language (EFL) textbooks, Ruegg and Brown analyzed vocabulary use in one text from 20 different books. Their findings show large amounts of high-frequency vocabulary used in the texts. When discussing language use in some of the textbooks for upper-intermediate learners, the two researchers claim that “it is highly likely that the vocabulary level of these books is pedagogically inappropriate” (2014, p. 17). This was due to the find of an overuse of vocabulary at the 1,000 frequency level in some textbooks for Upper-intermediate learners. The current study focuses on how general academic language is used in order to examine if it can be expected that L2 learners will be able to acquire these words implicitly.

Through the evidence provided in his research (see section 1.3.2), Hellekjær claims that elective English courses in Norwegian high schools do not challenge the students enough, blaming in part, textbooks in which the “texts are too often at a language and content level that provides little or no challenge for the students” (My translation from 2012a, p. 31). To the best of my knowledge, there are no studies conducted in relation to English course books written for Norwegian students that analyze general academic vocabulary use. It is my hope that the current study may provide new knowledge in relation to the use of general academic vocabulary in textbooks and if this vocabulary use provides the means for L2 implicit vocabulary acquisition through unassisted reading. I have not chosen to focus on elective English classes, but rather start with an analysis of textbooks for first year general studies students. I have done this because no such vocabulary analyses have been done for textbooks written specifically for Norwegian advanced L2 learners and it seemed appropriate to start with course materials written for the last obligatory English course.

1.3.2 Reading and L2 Vocabulary Acquisition

Many years of research show that there is a clear connection between an L2 learner’s vocabulary knowledge and their level of reading comprehension. As Cobb states, “there is

now widespread agreement among researchers that text comprehension depends heavily on detailed knowledge of most of the words in a text” (2007, p. 38). Cobb outlines some of the difficulties for L2 language acquisition in what he calls L2 language learners’ “lexical paradox”. This is the observation that “the words that occur in texts are mainly available for learning in texts themselves...[since] the lexis...of texts... in English...is far more extensive than the lexis of conversation...” (2007, p. 38).

After extensive research investigating how much vocabulary is needed for reading comprehension, there is a general consensus that for unassisted reading L2 learners should understand 98% of the words used in a text. A more conservative estimate of 95% word coverage should be seen as an absolute minimum (Laufer, 2010; Nation, 2013; N. Schmitt, Jiang, & Grabe, 2011). Expressed in more practical terms this means that learners should not be exposed to more than one unfamiliar word per 2-5 lines of written text, if they are to comprehend what they have read (Nation, 2013).

Nation (2006) conducted a pivotal study aimed at determining what vocabulary size L2 learners would need in order to reach 95% and 98% lexical coverage for reading comprehension of general written and oral English. Vocabulary size can be defined as “the number of words needed to meet a lexical coverage in various communicative contexts” (N. Schmitt, Cobb, Horst, & Schmitt, 2015, p. 2). Lexical coverage will be defined in this thesis as “what percentage of the vocabulary in a stretch of spoken or written discourse needs to be known by a learner in order for him or her to understand the discourse” (N. Schmitt et al., 2015, p. 2). The written corpus for this study contained English novels and newspapers. His findings showed that about 3,000 word families and proper nouns provided 95% lexical coverage, but to acquire the desired 98% coverage 8000-9000 families plus proper nouns were needed (Nation, 2006). In other words, Nation’s study revealed that L2 learners may need to have a very large vocabulary size i.e., “the number of words needed to meet a lexical coverage percentage in various communicative contexts”, in order to fully comprehend general written English (N. Schmitt et al., 2015, p. 2). It should be noted in this context that typically L2 English learners are expected to have a vocabulary size between 2,500 - 3,000 word families (N. Schmitt et al., 2015).

Glen Ole Hellekjær has conducted several studies in Norway related to students’ academic English reading proficiency (2009, 2012b). His studies focusing on high school seniors (2005, 2012b) have been quantitative studies that test reading comprehension using the International English Language Testing System (IELTS) and follow-up questions

regarding the personal reading habits of the participants. Because the test groups he used cannot be characterized as representative sample populations, one must be careful to generalize too much from the results. However, they do provide indications of trends among third year high school students' English reading comprehension and reading habits (Hellekjær, 2012b).

In discussions related to his studies, Hellekjær places focus on reading strategies that can improve reading comprehension. At the same time, he recognizes the importance of vocabulary knowledge in relation to reading ability (2012b). His comparative study of high school seniors in 2002 and 2011 also shows slight negative correlations between vocabulary knowledge and reading tests scores. The study showed that many students “lacked the ability to cope with unfamiliar words while they read”. Though these correlations are slight and cannot be used to form generalizable conclusions, it is interesting to note that among the test participants “the more often [they] say they have used word coping strategies, the lower their score on the reading comprehension test” (My translation 2012b, p. 164). These correlations then also raise the question of what type of vocabulary advanced L2 learners are exposed to during their studies, something the current thesis can help shed light upon.

Researchers today see the need for further studies related to the relationship between lexical coverage and vocabulary size in order to help provide a better understanding of how course materials may be written so that L2 learners may comprehend them (N. Schmitt et al., 2015). The current study uses a very small corpus so findings cannot be generalized to all textbooks, but at the same time, the corpus is directed at written input known to be used in English course classrooms in Norway. The qualitative aspects of the study, such as purposeful sampling of the populations and in-depth analyses, have allowed for focus to be placed on a more detailed discussion of AWL vocabulary. This has been done through the use of three different computer analyses. First an analysis to determine the use of AWL vocabulary and then two analyses to determine frequency levels of both the entire text and the AWL vocabulary found within topic related texts i.e., Range analyses (see section 3.5).

The corpus analyses also compare both tailored and authentic factual texts. As far as I know, this is something lacking in other studies. For the current study, the texts are defined as factual because they are either defined as such in the textbook index or they are non-fictional, topic specific texts linked to the textbook website. I will define authentic texts as “materials that were not originally developed for pedagogical purposes” (Richards, Schmidt, & Richards, 2002, p. 42). As Ellis and Shintani point out, “[t]here are marked differences in the

linguistic...and discourse features found in native-speaker corpora...and those found in language teaching textbooks” (2014, p. 166). Researchers debate both positive and negative aspects related to the use of authentic texts in classroom situations.

The aim for the investigation concerning authentic and tailored texts is to compare vocabulary use between them, but also to provide a clearer picture of what texts L2 learners might use in a classroom setting. The use of these texts will also provide a broader base for the analysis of topic related texts described in this thesis as narrow reading i.e., reading several texts related to one specific topic. It is important to stress that neither authentic or tailored texts are inherently good or bad. The key issue is that these texts need to help L2 learners achieve the goals set for them (Gilmore, 2007).

1.4 Research Questions

As outlined earlier, the current study will investigate the use of academic vocabulary in factual, textbook texts written for the target group. There are two main parts to this investigation, how is academic vocabulary used in textbooks written for English language students seeking higher education, and will this vocabulary usage provide the means for advanced L2 English students to acquire general academic vocabulary when reading factual texts. The following research questions will be investigated:

1. To what extent does the use of general academic vocabulary in factual, textbook texts provide the means for the implicit acquisition of this vocabulary during unassisted reading?
 - 1a. How is general academic vocabulary used within factual, textbook texts and across topic related texts?
 - 1b. To what extent does the use of glossaries in tailored texts assist advanced L2 English learners with the acquisition of general academic vocabulary during unassisted reading?

1.5 Outline of the Thesis

Through an in-depth analysis of AWL vocabulary use in factual texts related to commonly used textbooks for the target group, this study aims to investigate the use of academic vocabulary in textbooks while at the same time bring the term academic vocabulary into focus within a Norwegian context. At the heart of the study lies the discussion of implicit

vocabulary acquisition through unassisted reading. This vocabulary acquisition process will be presented and discussed in relation to usage-based theory and related theoretical hypotheses, such as the Frequency Hypothesis, the Noticing Hypothesis, and the Lexical Quality Hypothesis. Relevant research will also be presented and discussed throughout. The thesis is organized such that chapter two will provide a presentation of general academic vocabulary, reasoning behind the choice to operationalized the term using Coxhead's AWL, a brief discussion of Corpus Linguistics (CL) and a presentation of usage-based theory as the theoretical framework for the present study. In chapter three I will briefly describe the materials used and outline choices for research methods and provide an explanation of data collection methods. Findings from the text analysis will be presented in chapter four. Chapter five will provide a discussion of the findings before the concluding chapter presents summary remarks and explains implications for teaching and material design, as well as a discussion of further research areas in this field of enquiry.

2 Theoretical Background

The aim of this chapter is to present the term general academic vocabulary in more detail as well as to expound upon the reasoning behind the use of Coxhead's AWL in the current thesis. Following the discussion of academic vocabulary, a presentation of corpus linguistics will provide background information related to the data collection processes used in the current study. The usage-based theory guiding the current study will be presented and related to the research questions in the final sections.

2.1 General Academic Vocabulary

There is general agreement among linguists that L2 learners need to know large amounts of vocabulary in order to function well in academic settings. Setting vocabulary goals for L2 students pursuing academic studies at anywhere above 10,000 words is reasonable (Grabe, 2008, pp. 271, 279). Linguists also agree that students who wish to continue on to university studies must acquire proficiency in the use of general academic vocabulary (Coxhead, 2006; Grabe, 2008; Nation, 2013). The acquisition of general academic vocabulary can be difficult because they are often not salient i.e., they support the discussion, but are often not the main concept discussed and are seldom glossed (Coxhead, 2006; Flowerdew, 1993; Nation, 2013). The current study aims to investigate how general academic vocabulary is used in course materials written for Norwegian L2 learners and if this usage promotes the acquisition of these words.

It is important to keep in mind that the AWL used to define general academic vocabulary in the current study does not include discipline-specific academic vocabulary or GSL word families some see as academic in nature (Gardner & Davies, 2013). The AWL "is a good starting point because it has identified high-frequency, academic vocabulary, but it is not the sum total of academic vocabulary and learners will need to learn many words beyond that" (D. Schmitt, personal communication, Sept. 8, 2015). It is also important to note that other researchers have developed lists of general academic vocabulary, perhaps the most prominent of these today being the New Academic Vocabulary List developed by Gardner and Davies (2013). This list has not, as yet, been used to any great extent in research and will therefore not be discussed further in this thesis.

One aim of the current study is to place research regarding general academic vocabulary within a Norwegian context. The use of the term and debates about its relevance in English learning practices seem to be missing, despite the fact that there are clear indications from the national curriculum of the need to prepare Norwegian students for the use of English in institutions of higher education. In the outline for the purpose of teaching English in Norway the following is stated.

English is increasingly used in education and as a working language in many companies...To succeed in a world where English is used for international communication, it is necessary to be able to use the English language and to have knowledge of how it is used in different contexts

(Utdanningsdirektoratet, 2013)

These goals are also specified in the competence aims guiding both oral and written communication for the English subject. Students should be taught the ability to “express oneself fluently and coherently in a detailed and precise manner suited to the purpose and situation” (Utdanningsdirektoratet, 2013). Studies have shown that Norwegian students struggle with formal, academic language production (Hasselgren, 1994; Mahan & Brevik, 2013) and with reading comprehension related to academic texts (Hellekjær, 2008, 2012b).

2.1.1 The Academic Word List (AWL)

The following section contains a discussion of Coxhead’s development of the AWL, the debate connected with use of the AWL, and the decision to operationalize the term general academic vocabulary through the use of her vocabulary list.

In 1998, Coxhead compiled a list of academic words to help aid “teachers of [English for Academic Purposes] EAP courses set goals for their students’ vocabulary learning” (2011, p. 357). In order to develop the AWL, she compared written academic texts used in universities, from a wide range of subjects, and then compiled a corpus of 3.5 million words from 414 texts, covering the four subject disciplines: arts, commerce, law and science. Each subject discipline was divided into seven subject areas, such as education, accounting, and biology. Coxhead used the following set of criteria to determine which word families would be included on her academic word list (AWL). Word families included on the list had to

appear at least 100 times in the corpus, in at least 15 of the subject areas and over 10 times in each of the subject disciplines (2000). Coxhead decided not to include the 2,000 most frequent word families, as defined in the General Service List of English Words (GSL). There was a general assumption in this field of research that L2 learners would already know the GSL vocabulary (Nation, 2013). As a result of her research Coxhead then came up with a list of 570 word families described as academic vocabulary words prevalent in academic texts (2000).

The AWL has been used widely by researchers, materials developers, teachers and students alike since its publication (N. Schmitt, 2010). In recent years use of the AWL has been contested in several ways, resulting in a debate on the existence of general academic vocabulary (Hyland & Tse, 2007), questioning Coxhead's use of the GSL (Cobb, 2010; Gardner & Davies, 2013; Hyland & Tse, 2007), and the usefulness of the AWL in light of new frequency level developments (Cobb, 2010).

Hyland and Tse question the usefulness of a list of general academic terms. They dispute the need for the study of general academic vocabulary, contesting that “[i]t is by no means certain that there is a single literacy which university students need to acquire to participate in academic environments” (2007, p. 236). They advocate instead a need for students to study a “discipline-based lexical repertoire”. Hyland and Tse make a valid point that vocabulary words on the AWL can behave differently across disciplines. At the same time they acknowledge that Coxhead also “insist[s] that items should not be learnt out of context” (Hyland & Tse, 2007, p. 251). When addressing the issues raised here, Coxhead welcomes the discussion of placing AWL vocabulary in context and expresses the need for more research in line with the study conducted by Hyland and Tse (2011). The important factor is then to make sure that teachers understand that AWL vocabulary is not something to be taught as a list of decontextualized words, but must be used in close relation to texts written for academic purposes.

Perhaps a more important criticism of the AWL is related to Coxhead's decision to exclude GSL word families from the list. This was done in the assumption that the GSL (West, 1953) word families would be familiar to learners. The decision is something she herself describes as being controversial (2011). A major argument against the continued use of the GSL is the fact that the corpus forming the basis of the list stems largely from the early 1900's. Gardner and Davies (2013) claim “the GSL...is no longer an accurate reflection of high-frequency English”. In their criticism of the AWL, Gardner and Davies also point to the

fact that the GSL contains many “high-frequency academic words... like company, interest, business, market, account, capital, exchange and rate” (2013, p. 309). These words will not be categorized as general academic vocabulary if the AWL is used as a basis for such an analysis, such as is the case for the current study. Coxhead is aware of the difficulties related to her use of the GSL, but also points out that it has not been replaced and until this is “addressed in a careful and principled way”, the AWL should not be reworked (2011, p. 359).

When Nation’s (2006) study showed that learners might need vocabulary learning goals greater than the GSL plus AWL in order to comprehend general English, Cobb (2010) questioned the usefulness of the AWL (see section 2.2.2). He analyzed the GSL and the AWL with the BNC frequency levels developed by Nation and Beglar (2007). The study showed that GSL words were not as frequent as could be expected, with nearly 500 word families on the list outside of the first 2,000 BNC frequency levels. Nearly half of the 570 AWL word families were found at the first 2,000 BNC frequency levels showing that these two lists overlap greatly within the first 2,000 BNC frequency levels. However, when analyzing academic texts with the BNC and the GSL + AWL the later provided greater lexical coverage. This shows that, for academic texts “there is still room for an AWL” (Cobb, 2010, p. 193). Cobb proposed a modification of the AWL “within the BNC framework” (2010, p. 192). Coxhead confirms that the AWL falls roughly into the first 3000 frequency levels of the BNC. When questioned about Cobb’s proposal she felt this was an idea worth working towards (A. Coxhead, personal communication, June 4, 2015).

In a recent MA thesis written at the Concordia University in Montreal, a corpus of 15 university level economics textbooks was analyzed in order to develop a business English word list (Stella, 2015). During the process of conducting her study, Stella used different frequency lists to help remove general vocabulary from the corpus. She used the GSL + AWL, the New General Service List (NGSL) and the New Academic Word List (NAWL)¹, as well as the BNC-COCA frequency lists. The NGSL and the NAWL provided a slightly higher rate of lexical coverage than the older GSL and AWL lists, both close to 89%. Here lexical coverage refers to the percentage of commonly use vocabulary words within the corpus of business texts. However, the new lists contain nearly 1,200 more words, making the old lists “more cost-efficient for students because they have fewer words to learn” (Stella, 2015, p. 38). The BNC-COCA’s first 3,000 frequency levels provided the highest rate of lexical

¹ In 2013 Dr. Charles Browne, Dr. Brent Culligan and Joseph Phillips created a New General Service List (NGSL) and a New Academic Word List (NAWL) (2013).

coverage at 93%. Although Stella chose to use the BNC-COCA frequency lists for her study, she also claims that her results “challenge the hypothesis that the GSL and AWL would not be a good fit as core lists” (Stella, 2015, p. 36). Even though the GSL and AWL vocabulary lists are dated, they remain accurate tools for discovering the types of vocabulary they were meant to detect. Results from her study support my decision to use a combination of the AWL+GSL and the BNC-COCA for the current study.

I have chosen to use Coxhead’s AWL to operationalize academic vocabulary for several reasons. First of all, the AWL remains a widely cited and often used tool among researchers (N. Schmitt, 2010). It also continues to provide high levels of lexical coverage in academic texts and English language newspapers (Cobb, 2010; Nation, 2013; Stella, 2015). In order to compare my findings with previous studies it is important that I also define general academic vocabulary through the use of the AWL. Finally, the AWL is still used to analyze general academic vocabulary in the VP-classic program provided by Cobb (n.d.-b). This instrument was chosen because there is well-documented use of the site and it comes highly recommended by other researchers in the field (Coxhead, 2012; Nation, 2013; N. Schmitt, 2010).

2.2 Corpus Linguistics

Though the term corpus linguistics (CL) was not used extensively until the early 1980s, words have been indexed across texts since the 13th century. Today’s development of computer technology and internet access have taken the study of corpus linguistics to a new level (McCarthy, 2012). There is no explicit definition related to the use of CL in linguistics. Some researchers call it a discipline, others a methodology or a paradigm (Taylor, 2008). For the current study, I will focus on the use of CL as a method to obtain statistical data pertaining to vocabulary use.

One important aspect of the data collection process for my thesis is the use of computer programs to analyze vocabulary use. With the use of computer technology and the availability of a multitude of texts on the World Wide Web, corpora comprised of a billion words have been developed (McCarthy & O’Keeffe, 2012). The collection of these texts are used to provide linguists with “a means for the empirical analysis of language” which many researchers agree has led to better insight into the way language works (McCarthy, 2012, p.7). Many linguists will also agree with N. C. Ellis when he says, “Corpus linguistics ...[has] a

large role to play in identifying the linguistic constructions of most relevance to particular learners” (N. C. Ellis, 2012b, p. 204).

2.2.1 Corpora

The British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), are a collection of oral and written texts found in different authentic sources. COCA is made up of 450 million words that are “equally divided among spoken, fiction, popular magazines, newspapers, and academic texts” (Davies, 2012). Because of the recent collection of corpora texts, the number of texts, and its balance in text choices, the COCA has been described as the best corpora for general English in existence (D. Schmitt & Schmitt, 2012). The BNC “is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English, both spoken and written, from the late twentieth century” (Consortium, 2007). These two corpora, among others, have founded the basis for computer programs used by researchers to investigate vocabulary use and acquisition (Cobb, 2010).

Not all researchers agree with the use of CL, however. There remains a divide between linguists that place focus on observable data and those who prefer a theoretical approach to linguistic studies (Bonelli, 2012, p. 14). Aarts argues that, with regard to what encompasses CL, the focus is on “a *methodology* in linguistics”. As such, a corpus linguist will “use data from corpora in their claims about language” and therefore a corpus linguist can also be a theoretical linguist (2000, p. 7). Nevertheless, a debate continues among linguists regarding the study of corpus data in order to investigate language usage.

One often cited critic of CL is the American linguist Chomsky. He points to the importance of inquiry in all scientific research and not simply “taking photographs of what is happening in the world”. He continues, “You have to ask probing questions of nature. That’s what is called experimentation and then you get some answers that mean something” (Aarts, 2000, p. 6). Aarts points out that Chomsky has a point, as long as CL is only used to observe, f. ex. word-frequency, “without couching those data within the framework of meaningful questions about the structure or usage of the language being studied” (2000, p. 7). Most corpus linguists today will agree that it is important to study language within its context and when this is done it can be very useful to “count things”, but Aarts also calls for the use of “qualitative data that corpora can furnish” (2000, p. 8).

It is the aim of the current study to both “ask probing questions” and let corpora inform the mixed methods study of academic vocabulary. As such, it is my goal that the methodology applied to the study will result in much more than a mere “picture of events”.

2.2.2 Counting Words

One important yet difficult task when using corpora for analysis is for program developers to determine what counts as a word. In CL, the terms *token*, *type*, *lemma* and *word family* are used to differentiate between items that are counted during computer analyses. The following section provides definitions of the terms and how they may be applied in corpus studies.

Tokens refer to every word form found in a written or oral text. If the same word form, i.e. *language* occurs several times within a text each form is counted as a separate entity. The term *types* refers to a gathering of tokens so that if one word is written several times in a text these are counted in one group and referred to as one word type. (Nation, 2013). If the word *language* is used seven times within one text then this will form the word type group of *language*. Likewise, if the word *languages* appears five times in the same text, this will be counted as one word type group as well.

In corpus studies, words that are closely related are often counted together. This can be done by the use of *lemmas* or *word families*. *Lemmas* refer to “a head word and its inflected forms and reduced forms” (Nation, 2013, p. 10). Usually all of the items in a lemma are all the same parts of speech (Francis and Kucera (1982) in Nation, 2013). Counting similar forms as two different words is often times not helpful in text analyses because we can assume that an L2 learner will know the meaning of similar word types, such as *language*, and *languages*. It can be assumed that these learners know regular patterns related to plural forms of nouns. In this manner, if the base word *language* is known, the plural form *languages* will have a low learning burden. The learning burden of lexis refers simply to “the amount of effort required to learn [an item]” (Nation, 2013, p. 10).

Word families represent “a headword, its inflected forms, and ...closely related derived forms” (Nation, 2013, p. 11). Nation argues that affixes, such as, *-ly*, *-ness*, and *un-* “greatly reduce the learning burden of derived forms containing known base forms, but he also acknowledges that not all learners will necessarily know all of the derived forms in a word family. The learning burden of derivatives can be discussed. Studies have shown that L2 learners learn derived forms much later than inflected forms (Gardner & Davies, 2013;

Nation, 2013). “What might be a sensible word family for one learner may be beyond another learner’s present level of proficiency” (2013, p. 11). In other words, the use of word families in corpus analyses assumes the L2 learners have knowledge of all the inflected and derived forms of a headword. This might not be the case for less proficient learners.

In the text analyses for this study, word families form the basis for the counting of words in the computer programs that are used. Therefore, the analyses for the current study are based on tokens, types and word families. This choice was made because of my reliance on computer programs and vocabulary word lists that use word family divisions. The investigation of general academic vocabulary use is related to advanced L2 learners and, as such, it can be expected that many derivatives of words will have a relatively low learning burden for these learners. I am, however, aware that the use of word families can pose difficulties also for advanced L2 learners. Ways in which tokens, types and word families apply to the current study are outlined in section 3.5.1.

2.2.3 Using CL to Determine Frequency Levels

Vocabulary acquisition for L2 learners is a complex and time consuming process. With this in mind, one question vocabulary researchers have been seeking answers to is what vocabulary L2 learners should learn first. CL has informed this investigation with use of such word lists and corpora as the General Service List (GSL), the AWL, the BNC and, more recently, the COCA. These corpora have been used to create levels or bands of the 1000 most to least frequently used English words, here referred to as frequency levels. Researchers, teachers and students have been able to use this information in order to create lists of high-frequency words L2 learners should learn first (Cobb, 2010, n.d.-a).

For this investigation, one on-line site has been used to conduct several different analyses of textbook texts, part of which are related to gathering frequency data. The aim of the study is not to simply collect this information, but to let this data inform a discussion as to the possibilities that academic vocabulary is used in these texts in such a manner that may provide the means for L2 vocabulary acquisition of academic vocabulary. A further discussion of the use of corpora and computer programs for the current study is given in section 3.5.

2.3 Usage-based Theory

In usage-based theory, language is seen as a “complex dynamic system” a system in which “language emerges through use and changes continually because of interactions at all levels” (N. Schmitt & Verspoor, 2013, p. 353). The complexities of the cognitive processes involved in implicit vocabulary acquisition during unassisted reading and the use of cognitive linguistics have shaped the language acquisition theory used for this thesis.

Usage-based approaches may help broaden our understanding of the complex processes involved in L2 vocabulary acquisition by investigating patterns in language usage and then applying these to more complex questions, such as if reading factual texts may promote implicit vocabulary acquisition. According to N.C. Ellis, the main “motivation of usage-based approaches is [to] bring together linguistic form, learner cognition, and usage” (N. C. Ellis, 2015, p. 263). For the current study, the linguistic form of general academic vocabulary will be examined in order to understand if the use of this form in textbook texts may assist advanced L2 learners acquire more formal vocabulary implicitly.

In usage-based models of language “the linguistic system is fundamentally grounded in ‘usage events’: instances of a speaker’s producing and understanding language” (Kemmer & Barlow, 2000, p. viii). This means that the linguistic system used in communication contains use of both grammar and lexis that may influence those exposed to the discourse. As such, “language productions are not only products of the speaker’s [linguistic] system, but...also provide input for other speaker’s systems... Thus, usage events play a double role in the system: they both result from, and also shape, the linguistic system itself in a kind of feedback loop” (Kemmer & Barlow, 2000, p. ix). For the current study, this means that the vocabulary used in the textbook texts is perhaps influenced by the fact that they are written for advanced L2 learners and at the same time, the L2 learners reading these texts may be influenced by the vocabulary usage prevalent in the texts.

Verspoor and Schmitt apply usage-based theories to L2 vocabulary acquisition (2013). They explain this view of L2 vocabulary acquisition as follows:

[It] is an assembly of meaningful, symbolic units, which can be words, formulaic sequences, idioms, or longer syntactic constructions. They are learned through exposure in a bottom up process—with the help of some basic cognitive abilities such as association, categorization and schematization. The more frequently a unit is heard or used and the more

meaningful clues the learner has, the more chance initial form-meaning links are made and the more chance the form will be used in conventionalized contexts (2013, p. 356).

Focus is placed on the importance of contextual settings and repeated exposure to individual language items. There is also recognition for and understanding of language as something that is constantly changing through the power of the speakers themselves. In a globalized world, this aspect of the English language seems particularly important to acknowledge.

Langacker defines association as a “phenomenon...in which one kind of experience is able to evoke another”. Association as such, is directly related to symbolization “the association of conceptualizations with the mental representations of observable entities such as...written marks” (Langacker, 2000, p. 5). In vocabulary acquisition, symbolization can refer to the creation of a form-meaning association i.e., a symbolic unit. This form-meaning association is expanded upon with repeated exposures (N. Schmitt & Verspoor, 2013, p. 354). Abstraction is “the emergence of a structure through reinforcement of the commonality inherent in multiple experiences”. Schematization is a form of abstraction “involving our capacity to operate at varying levels of ‘granularity’” (Langacker, 2000). In other words, after experiencing a usage event, such as a form-meaning association in vocabulary, the learner can begin to group specific qualities of one word with the same qualities of other words and thus group them into categories. An example could be the word *table*. After many experience of the word it is possible for L2 learners to find comparable attributes that form a category representing a piece of furniture, normally with four legs that can be use to put things on. This means that the symbolic unit has become abstract; carrying more meaning than the initial unit and as such has been categorized or grouped. In vocabulary acquisition, this process can, among other things, relate to the creation of form-meaning receptive word knowledge which is an important part of the current investigation related to both implicit vocabulary acquisition and the discussion of word knowledge (see section 2.41).

The remaining discussion of theory is centered on a more detailed account of the theoretical framework in which specific hypotheses have been used to provide a theoretical basis for the different types of analyses conducted in the current study. Following a brief discussion of each hypothesis, relevant research will be examined. Finally, both the theoretical aspects and research will be related to the research questions for this thesis.

2.4 Relevant Hypotheses

There is no one theory describing the L2 vocabulary acquisition process. I have related different SLA hypotheses to implicit vocabulary acquisition through reading and placed these within a usage-based framework in order to emphasize the study's focus on an assessment of vocabulary use. The hypothesis discussed are the Frequency Hypothesis, the Noticing Hypothesis and the Lexical Quality Hypothesis.

The L2 acquisition hypotheses presented here are directly related to my research questions that structure the current investigation of general academic vocabulary usage in English course materials. Usage-based theory describes vocabulary acquisition through the development of form-meaning association provided in usage events i.e., occurrences of speaker production and understanding. As such, central elements of the implicit vocabulary acquisition process relevant to the thesis are input, frequency and attention (N. Schmitt & Verspoor, 2013).

2.4.1 The Role of Input

Input plays a central role in all second language acquisition (SLA) theory because there is acceptance for the “general assumption that no learning can take place unless learners have access to input” (R. Ellis & Shintani, 2014, p. 174). It is assumed, and many much empirical evidence supports the idea (Elgort & Warren, 2014; Kang, 2015) that L2 learners can acquire vocabulary while reading texts on their own i.e., implicit vocabulary acquisition through unassisted reading.

The implicit acquisition of language features is understood to be “learning that takes place without either intentionality or awareness” (R. Ellis, 2008, p. 7). However, it should be noted that not all researchers agree that language acquisition can be achieved without some degree of consciousness (see section 2.4.3). At the other end of an L2 acquisition continuum is explicit learning, which refers to learning as “a conscious process and is likely to be intentional” (R. Ellis, 2008, p. 7). In the current study I will use the terms as defined here.

Determining the exact nature of explicit and implicit learning has been at the heart of many theoretical debates in the field of L2 acquisition (N. C. Ellis, 1994; R. Ellis, 2008; Krashen, 1981). Perhaps the greatest proponent for L2 implicit learning through reading is

Krashen. His Input Hypothesis regarding L2 acquisition² claims that “we acquire...only when we understand language that contains structure that is ‘a little beyond’ where we are now..., [i.e.] $i + 1$ ” (1982, p. 21). He claims that academic vocabulary will also be acquired through reading as long as students receive “more comprehensive input [$i + 1$]”³. Krashen is very specific about this in relation to the acquisition of academic vocabulary, claiming that “if it is academic vocabulary it will be acquired through reading.” He thinks researchers who claim the need for the explicit learning of academic vocabulary are mistaken. He goes so far as to claim that teaching vocabulary “is an approach that has never worked” (2013, p. 28). Krashen’s reasoning behind this claim is that “direct instruction cannot deal with the size and complexity of vocabulary. There are simply too many words to be acquired” (2013, p. 33). He does concede however that the “process is gradual and quite a bit of reading is required to build a sizable vocabulary” (Krashen, 2013, p. 29). Though there are many who disagree with Krashen’s Input Hypothesis, it is difficult to discuss implicit vocabulary acquisition through reading without mentioning the hypothesis.

While most linguists agree that vocabulary cannot be acquired solely through explicit learning, many find the explicit instruction of vocabulary necessary for the same reasons Krashen finds this impossible. Norbert Schmitt claims that a more “proactive, principled approach needs to be taken in promoting vocabulary learning which includes both explicit teaching and exposure to large amounts of language input, especially through extensive reading...” (N. Schmitt, 2010, p. 8). According to Nation, L2 learners “should not rely solely on incidental vocabulary learning from context”. Many agree with Nations’ view that implicit and direct vocabulary learning are “complementary activities” (2013, p. 357). Nation claims that “deliberate vocabulary learning...is not only efficient but effective, in that such knowledge can be retained...and involves the implicit knowledge which is essential for normal language use” (2013, p. 217). Here focus is on the time factor, since such a large amount of vocabulary is necessary in order to acquire a small amount of mid-frequency vocabulary implicitly (see section 2.3.4).

² It should be noted that by acquisition Krashen refers to what many would call implicit learning, i.e. learning that does not involve the “explicit teaching of rules” (Krashen, 1981, p. 1).

³ Though this is not directly related to the current discussion, it should be noted that some researchers attempt to compare the input hypothesis to Vygotsky’s ZPD. Dunn and Lantolf claim that this comparison is “unproductive”. Their argumentation is focused on differences between the input hypothesis and the ZPD in concepts related to the learner and his/her learning process. In Krashen’s hypothesis the learner is “a passive body” compared to Vygotsky’s acquiring through “collaborative activity”. They also point towards differences in learner autonomy with Krashen’s autonomous learner versus “personal ability co-constructed through activity with other people and artifacts” in the ZPD (From Dunn and Lantolf 1998 in Lantolf & Thorne, 2006, p. 273).

2.4.1.1 Support from research

While there is a multitude of research to support the importance of input on L2 vocabulary acquisition, I have chosen a more in-depth discussion of a recent study with relevance to different aspects of the theoretical framework in my research. The findings provide recent support for the role written input can play in the vocabulary acquisition through the study of implicit vocabulary acquisition in relation to both frequency and noticing. In order to provide a coherent presentation of the study I will present the findings relevant to the current study here even though they are interrelated with things that will also be discussed later in relation to my own findings.

Elgort and Warren (2014) have shown that L2 learners acquire both explicit and tacit word knowledge i.e., knowledge difficult to verbalize, during unassisted reading. Before further discussion of their study, it is necessary to provide a brief description of explicit knowledge. The facets of explicit knowledge in focus for the current study include knowledge in which the learner has a conscious awareness of linguistic forms, can be reported and is learnable at any age (R. Ellis & Shintani, 2014, p. 13).

In a quantitative study of 48 university students in New Zealand with English as their L2, Elgort and Warren examined both explicit and tacit knowledge gained from reading several chapters in a non-fiction book about economics (Elgort & Warren, 2014). Explicit knowledge, unlike explicit learning, refers in this study to form-meaning vocabulary knowledge gained through implicit vocabulary acquisition during unassisted reading (see section 2.4.2.2). Each participant read the text in their spare time, were told to read for meaning, were also instructed to only read the text once and told that they should not use a dictionary to look up words. After reading each chapter, participants were given comprehension tests. They were given pretests to measure vocabulary size and language proficiency, as well as posttests of word meaning (Elgort & Warren, 2014).

It should be noted that Elgort and Warren use the term incidental L2 learning in their study. I have chosen to use the term implicit L2 learning in my discussion of the study because of my use of the definition provided by Ellis and Shintani (2014). They define implicit learning as “[l]earning that takes place without ...intentionality or awareness”. Research involved in investigating this type of learning he defines as “exposing learners to input data, which they are asked to process for meaning, and then investigate (without warning) whether they have acquired any L2 linguistic properties as a result of the exposure” (R. Ellis & Shintani, 2014, p. 338). Here they distinguish between incidental L2 learning

research as “investigated by giving learners a task that focuses their attention on one aspect of the L2 and, without pre-warning, testing them on some other feature” (2014, p. 338). I will not go into a further discussion of the differences between incidental and implicit L2 learning, but wish to reiterate my use of the term implicit L2 learning based on Ellis and Shintani (2013).

Elgort and Warren (2014) showed that unassisted reading provided several positive effects related to gains in explicit vocabulary knowledge acquired through implicit learning during unassisted reading. The average amount of word meaning learned implicitly was 10 out of 48 items, which is a rate comparable to previous studies (Elgort & Warren, 2014). The amount of vocabulary gained implicitly during reading was shown to be closely related to L2 learner vocabulary knowledge. “[R]eaders with higher comprehension scores were better able to take advantage of multiple occurrences...More advanced participants were more likely to learn the meanings of critical items after fewer encounters compared to less advanced participants...[and] contextual word learning progressed faster for readers who reported greater use of vocabulary learning strategies” (Elgort & Warren, 2014, pp. 380-381).

Student participants who had low comprehension of the text were “less likely to learn the meanings of the new vocabulary items even if they occurred multiple times in the text” (Elgort & Warren, 2014, pp. 395-396). For those L2 learners with lower proficiency levels, learning the vocabulary items from context was only successful when the words occurred within the same chapter. The study supports previous studies showing that recency i.e., the time in between word exposures, may also play an important part in vocabulary acquisition (Elgort & Warren, 2014). Their results showed that few vocabulary items were gained with under 12 repetitions.

In relation to the current study, these findings support theoretical descriptions of the theoretical framework supporting my research. The findings are related to the importance of input and frequency. The study also supports the recognition that implicit vocabulary acquisition is a slow process that demands large amounts of text (Cobb, 2007; Krashen, 2013; Nation, 2013). Finally, the findings provide support for the Lexical Quality Hypothesis that claims, among other things that more proficient readers will learn vocabulary more easily (see section 2.4.4).

In general, the study showed “that number of encounters with a new word in reading is a major predictor of gain in explicit word knowledge. However ...contextual learning is more

problematic for less proficient L2 readers. Thus...reading needs to be supplemented with deliberate word learning and vocabulary learning strategies” (Elgort & Warren, 2014, p. 397). As such, the findings suggest that L2 learners with a small vocabulary size and less advantages reading skills gain less vocabulary through unassisted reading. The researchers conclude that explicit vocabulary instruction should accompany reading. It is this discussion of implicit and explicit acquisition of general academic vocabulary that is in focus for this study.

2.4.2 The Frequency Hypothesis

The Frequency Hypothesis was coined by Hatch and Wagner Gough and “stated that the order of L2 acquisition is determined by the frequency with which different linguistic items occur in input” (Hatch & Wagner-Gough (1976) in R. Ellis, 2008, pp. 241-242). Ellis holds that “[f]requency is a key determinant of acquisition because ‘rules’ of language, at all levels of analysis...are structural regularities which emerge from learners’ lifetime analysis of the distributional characteristics of the language input” (N. C. Ellis, 2012b, p. 196). In other words, it is through exposure to a target language that L2 learners acquire “rules” of linguistic features, and not necessarily by memorizing rules governed by others. The more often learners are exposed to certain language features, the more likely they are to acquire knowledge of these features. According to the Frequency Hypothesis, learning is exemplar-based rather than rule-based (R. Ellis & Shintani, 2014). This corresponds well with usage-based theory of L2 vocabulary acquisition in which symbolic units are categorized and placed into different schema after more exposure to the symbolic unit leads the learner to gain abstract knowledge of this unit (N. Schmitt & Verspoor, 2013).

The Frequency Hypothesis applies to vocabulary acquisition in the same way it applies to the acquisition of grammatical patterns. “[L]earners learn words that occur frequently in the input before those that occur less frequently” (R. Ellis & Shintani, 2014, p. 175). Frequency in terms of L2 language acquisition is often referred to as token or type frequency. Token frequency is related to the number of occurrences of certain linguistic forms within input and type frequency concerns “items that occur in a slot in a construction” (R. Ellis & Shintani, 2014, p. 176). The analyses conducted in the current study focus on token frequency. These specifications of frequency should not be confused with word types and tokens, however (see section 2.4.1).

One implication of the Frequency Hypothesis is the need for L2 learners to access large amounts of input in order to “fine-tune their developing knowledge” (R. Ellis & Shintani, 2014, p. 176). One question vocabulary researchers have asked is related to how much input L2 learners need in order to acquire different types of word knowledge. With this in mind, linguistic research is increasingly returning to the importance of frequency. Ellis points towards psycholinguistic research that has, during the past 50 years, “demonstrated language processing to be exquisitely sensitive to usage frequency at all levels of language representation” (N. C. Ellis, 2002). Further, he argues that “frequency effects [provide]...compelling evidence for usage-based models of language acquisition which emphasize the role of input” (2012b, p. 199). With regard to this study, written input and vocabulary acquisition through reading are in focus. The frequency of AWL word families experienced during reading guides the investigation of written input provided in factual, textbook texts, as this is an area of research few have investigated previously in Norway.

2.4.2.1 Zipf's Law

The phenomena uncovered in the above mentioned study can be explained through Zipf's law, which describes, through a mathematical equation, “the nature of vocabulary use” among other things (Nation, 2013, p. 33). More precisely the law has shown that, “the most frequent word [in a text occurs] approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, etc.” (N. C. Ellis, 2015, p. 262). The implication of Zipf's law is that the implicit acquisition of less frequent English vocabulary words through unassisted reading requires a lot of reading because it will take large amounts of input for L2 learners to encounter less frequently used vocabulary enough times. In fact “it is not unusual to find lots of words occurring once in course books written for learners of English” (Nation, 2013, p. 33).

In the current study, AWL vocabulary has been analyzed both in terms of percentage of use and number of repetitions in and across texts. An investigation of AWL vocabulary occurring only once has also been included.

2.4.2.2 Word Knowledge

As mentioned in the introduction, word knowledge is a complex concept. Among linguists “there is no agreed terminal stage for knowledge of a word” (R. Ellis, 2008, p. 99). In fact,

today more and more attention is given to the acquisition of L2 vocabulary in chunks of words, or collocations. There is not agree definition for collocations, in this study I will define the term broadly as “words that tend to cluster together in the same textual environment” (Read, 2000, p. 232). It is becoming apparent that words are largely learned in groups (Nation, 2013; N. Schmitt, 2010). Though the study of collocations is an important development in applied linguistics, it is unfortunately beyond the scope of this thesis. Because words work together, “there are many things to know about any particular word and there are many degrees of knowing” (Nation, 2013, p. 44).

Due to the complexity of this concept, I will provide only a brief description of word knowledge. Nation outlines three types of word knowledge, form, meaning and use. There are many forms related to one word including aspects of phonology as well as written forms. Meaning knowledge includes basic knowledge of one way a word can be used to a more abstract knowledge of how the word functions in context. Word knowledge related to use include knowledge of how to use words correctly in grammatically, what other words they work with (collocations) and when not to use the word. Each of the three parts are related to both receptive and productive knowledge. Productive knowledge is more difficult to obtain than receptive knowledge because this type of word knowledge requires good enough understanding of the term to be able to reproduce the word with the correct form and the correct meaning, in the right context. Receptive word knowledge in the most basic sense includes recognizing the form of a word and being able to connect the correct meaning to the word (Nation, 2013, p. 48).

During reading, L2 learners must have form-meaning, receptive knowledge of vocabulary items in order to have a most basic form of comprehension (Brown, 2015). “[R]eceptive vocabulary use involves perceiving the form of a word while listening or reading and retrieving its meaning” (Nation, 2013, p. 47) This process of receptive knowledge means that the learner comprehends the word when seen, but does not necessarily have appropriate word knowledge to use the word productively in written or spoken discourse. “Productive vocabulary use involves...producing the appropriate spoken or written word form” (Nation, 2013, p. 47).

2.4.2.3 Support in research

In an early Swedish study of L2 vocabulary acquisition, the productive vocabulary of 11-year-old L2 English learners’ was studied over time. This longitudinal study showed that a

majority of the productive increase in vocabulary (over 60%) related to words found in their textbook (Palmberg (1987) in R. Ellis, 2008). Palmberg's study shows the effects of written input on vocabulary learning, but it did not include an analysis of frequency, but many other studies have. The question of how many repetitions of a word are needed to facilitate the implicit L2 acquisition of a word has been examined extensively.

Most researchers in the field today will agree that there should be a minimum of six occurrence for L2 learners to gain this knowledge, while ten to twenty repetitions is preferable (Nation, 2013; N. Schmitt, 2010). New advances in technology have made it possible for researchers to monitor the eye movement of readers to determine how much time is spent reading the different lines of texts; the assumption is that less time spent reading new words means that the readers have acquired form-meaning knowledge of these words (Pellicer-Sánchez, 2015). A recent quantitative study using eye-tracking methods and follow-up vocabulary knowledge tests has shown that “[a]fter eight exposures, L2 readers recognized the form and the meaning of 86% and 75% of the target non-words, respectively” (Pellicer-Sánchez, 2015, p. 1). Pellicer-Sánchez tested vocabulary knowledge gained incidentally from reading. The participants were not told they would be given vocabulary tests after reading a short story containing nonwords, therefore the study is considered a measure of incidental vocabulary acquisition. For this study 23 of the participants had English as an L2. They were undergraduate students attending a institution for higher education in the UK. The earlier mentioned study conducted by Elgort and Warren (2014) showed few participants gained word meaning knowledge of words that were repeated under twelve times. For the current study, I have chosen to use the minimum of six AWL word family occurrences within and across texts for the text analyses. This has been done in order to allow for a discussion of rates that also include the minimum requirements researchers agree to.

Another question under investigation by vocabulary researchers has been related to the amount of input necessary to provide enough repetitions needed for implicit vocabulary acquisition through unassisted reading. Cobb (2007) investigated what Krashen's reference to “quite a bit of reading” actually could mean for L2 learners. In his quantitative study, Cobb researched the in-text frequency of word families with BNC frequency levels at the first three thousand levels. He compiled a 517,000 token corpus of fiction, press writing and academic writing taken from the Brown corpus and searched for repetitions of ten word families from each frequency level. He found that nearly all of the 1,000 level words were repeated more than six times in each corpora, but only half of the 3,000 level words were repeated six or

more times (Cobb, 2007). He found that the rate of recycling after the first 2,000 most frequently used word families dropped drastically. The implication of this is that more advanced English L2 learners will need to read much more in order to encounter enough repetitions in context to gain form-meaning knowledge of words occurring at the 3,000 frequency levels and above.

2.4.3 The Role of Noticing

While “[f]requency is important in tuning the system, ...it is by no means the only factor that counts in acquisition” (N. C. Ellis, 2012a, p. 26). The concepts of attention, awareness and noticing, are also important in a theoretical discussion of SLA pertaining to this thesis. The following section will present a brief overview of the theoretical discussion centered around these terms, as well as how they relate to what is often referred to as the Noticing Hypothesis. Implications of this theoretical discussion for L2 vocabulary acquisition and its relevance to the current study will also be presented. The discussion will be supported by relevant research.

2.4.3.1 The Noticing Hypothesis

As mentioned earlier, implicit learning implies learning without intentionality or awareness (R. Ellis, 2008). Schmidt’s Noticing Hypothesis questions the assumption that language can be acquired without conscious effort. The hypothesis poses that “input does not become intake for language learning unless it is noticed”. The behind Schmidt’s hypothesis is that “SLA is largely driven by what learners pay attention to and become aware of in target language input” (Schmidt, 2010, p. 721). His hypothesis “postulates that attention controls access to awareness and is responsible for noticing” (Leow, 2013, p. 42). *Attention* in SLA refers to the need for recognition of a linguistic feature in the L2 input to be noticed at least at a low level of awareness (Leow, 2013). Attention has “four main characteristics: its capacity is limited, it is selective, it is voluntary, and it controls access to consciousness” (Leow, 2013, p. 44). Schmidt defines two levels of *awareness* i.e., *awareness at the level of noticing* and *awareness at the level of understanding* (1995). He defines awareness at the level of noticing as “conscious registration of the occurrence of some event” compared to awareness at the level of understanding “impl[y]ing recognition of a general principle, rule or pattern” (1995, p. 29). He argues that “[l]earning...is...largely, and perhaps exclusively a side effect of attended processes” (Schmidt, 2001, p. 29). However, he does acknowledge implicit learning

as is shown in research in which “the result of allocating attention to input results in more learning than can be reported verbally by learners” (2001, p. 4) i.e., they demonstrate tacit knowledge.

Placing the two terms of *noticing* and *understanding* into L2 a discussion of vocabulary acquisition, Schmidt defines the “conscious registration of the form” as *noticing*, while *understanding* refers to “[k]nowing the meaning of the word and knowing its syntactic privileges of occurrence (other than collocations and fixed expressions)” (1995, p. 29). As such, the use of glossary items that are in focus for the current study may both aid *noticing* and lead the L2 learner towards *understanding*.

2.4.3.2 *Glossing as a function of awareness*

Nation defines a gloss as “a brief definition or synonym, either in L1 or L2, which is provided with the text”. Glossaries may be used to help L2 learners in several ways. They may help students read texts that otherwise would be too difficult and they provide definitions for words that may not be guessed properly. In addition glossing “provides minimum interruption of the reading process...and draws attention to words [which] may encourage learning” (Nation, 2013, p. 238). Glossing is one way of enriching written input. The use of glossaries can help learners “construct the form-meaning mapping that is central to L2 acquisition” (R. Ellis & Shintani, 2014, p. 190).

There are certain considerations to be taken when it comes to the development of glossaries that can best help promote L2 vocabulary acquisition. These are related to where glossaries appear on the written page, if they should be translations or L2 definitions and how many words in a text are glossed. Studies show that learners have a tendency to prefer glosses in the margins of the text, and as research in the area does not find differences between glosses in the margin or at the end of the text, Nation recommends following the will of the reader (Nation, 2013; Rott, 2002). With regard to the amount of glossing in a text, Nation recommends around 3% but no more than 5% of the text should be glossed, though there is no experimental or observational data on this topic. Research has shown no difference between the use of L1 and L2 glossaries in relation to effects on comprehension or acquisition. Vocabulary frequency and density data should also be taken into consideration when developing glosses. Nation poses “a useful rule [to follow] would be to gloss mid-frequency words” (2013, p. 242). “Most studies have found that glossing has a positive effect on

vocabulary learning (Nation, 2013, p. 244). At the same time, glossing must be in relation to the L2 learners' vocabulary size (Rott, 2002).

There are 15 of the 21 texts analyzed in the current study that use glossaries, these are all tailored texts. Since all three of the textbooks for this study will have used marginal glossaries with L1 translations of the words, a discussion of these aspects will not be in focus. Instead, focus has been placed on glossary text coverage (how much of the text is glossed), AWL glossary coverage (how many of the AWL word families in the text are glossed) and frequency levels of glossed terms. The current study has investigated to what extent the use of glossaries in factual, textbook texts may help assist advanced L2 English learners with the acquisition of general academic vocabulary. To determine facets of glossing that should be included in the analysis, it has been necessary to define glossing and to access aspects of glossing researchers recommend in order for glossed terms to best aid L2 learners with vocabulary acquisition of general academic words.

2.4.4 The Lexical Quality Hypothesis

Researchers have found clear connections between an L2 learner's vocabulary knowledge and their level of reading comprehension (Cobb, 2007; Laufer, 2010; Nation, 2013). These gains in understanding “[have] been guided by specific problems and flexible frameworks more than by the testing of precise theories” (C. C. Perfetti & Stafura, 2014, p. 22). One such framework has been proposed by Perfetti and Hart, which they call the Lexical Quality Hypothesis (2002). It models a causal relationship between L1 vocabulary proficiency and reading comprehension. They argue that “skill in reading comprehension rests to a considerable extent on knowledge of words” (2002, p. 189). Their hypothesis is depicted in a model which describes the positive effect both have on each other. In other words, the more vocabulary a person knows, the more competent they will be as readers, and the more competent they are as readers, the more vocabulary knowledge learners will acquire. Their model is based on L1 vocabulary proficiency (C. A. Perfetti & Hart, 2002).

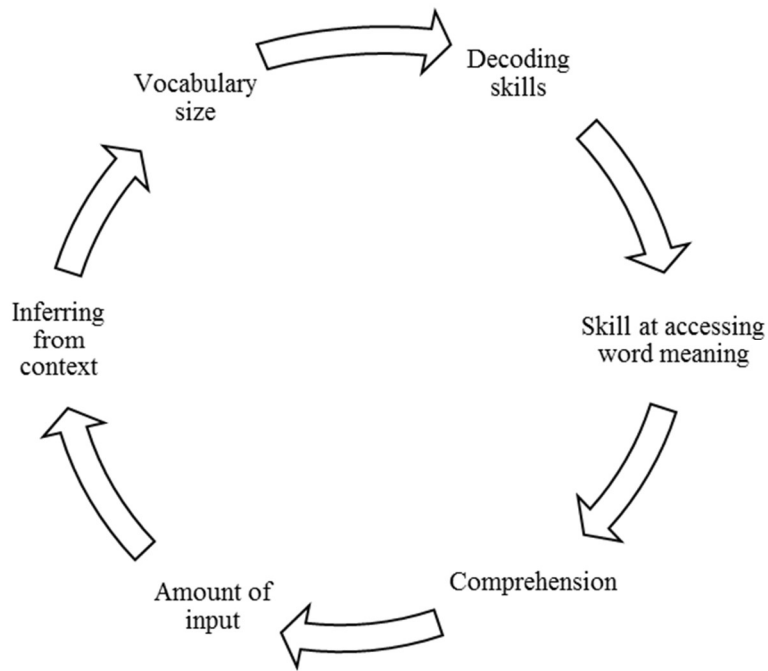


Figure 1. *An expanded model of the Lexical Quality Hypothesis*

Perfetti expanded the original model in 2010 to include skills in decoding and accessing word meaning (Perfetti (2010) in Nation, 2013). Nation expands the model, in a discussion of L2 vocabulary acquisition, to also include “skill at inferring from context and vocabulary size”. As such, it can be said that this framework provides a simplified model that represents “the skills, knowledge and experience that are essential for vocabulary growth through reading, and there is research support for the parts of the model” (Nation, 2013, pp. 350, 351). Again, the model shows that an increase in vocabulary learning skills will lead to an increase in reading comprehension, which will in turn lead to an increase in vocabulary learning skills. Learner who have good skills at inferring meaning from content develop larger vocabulary size and the larger a vocabulary size L2 learners have. Having a large vocabulary size supports decoding skills i.e., phonological and orthographic knowledge and vice versa (C. Perfetti, 2010). In other words, a learners knowledge of oral and written features connected to language usage events will aid their ability to learn new words and learning new words will provide possibilities to increase word knowledge i.e., vocabulary size. It should be noted that the model is a simplification of this process in that only written texts are considered, the relationships between the different parts are oversimplified as being linear relationships, and the model also omits important factors such as motivation and word knowledge (Nation, 2013).

When applying the framework outlined in the Lexical Quality Hypothesis to this study several assumptions are made. There is an assumption that L2 learners should “know” between 98% and 95% of the vocabulary present in a text in order to properly comprehend the content of the text. Another assumption is that factual, textbook texts will provide more exposure to general academic vocabulary than works of fiction. It is assumed that with six repetitions or more L2 learners may implicitly acquire form-meaning, receptive knowledge of these vocabulary words; this is accomplished largely by inferring meaning from the context. There is also an assumption that the L2 learners using the analyzed textbooks should have adequate vocabulary knowledge of English vocabulary at the first 2,000 BNC/COCA frequency levels. For the current study, it is also assumed that high-frequency vocabulary is easier to learn than vocabulary at lower frequency levels, largely due to the Frequency Hypothesis, though there are differences between L1 and L2 learners in this area. It is especially more likely for L2 learners to have greater knowledge of technical or discipline-specific vocabulary (N. Schmitt & Verspoor, 2013). The current study focuses on general academic vocabulary so the implications are perhaps not as strong, but are something to be aware of. The same is true with regard to glossing and L1 to L2 transfer, especially when glossary terms may help L2 learners acquire understanding of the vocabulary more rapidly if the word is known well in the L1 and is easily translated (N. Schmitt & Verspoor, 2013). However, a closer investigation of L1 to L2 transfer is out of the scope of the current study and will not be discussed further.

2.4.4.1 Reading strategies

While the hypothesis shows the complementary relationship between reading comprehension and vocabulary knowledge, vocabulary acquisition is also dependent on reading strategies. Different ways of reading texts are especially prevalent within a classroom setting. Focus has been placed on narrow reading for the analyses in the current study. This has been done in order to try and better simulate actual reading practices that might be used in the classroom. Research has shown that these different approaches to reading a text may promote different types of vocabulary acquisition (Nation, 2013). *Narrow reading* refers to “reading within a very narrowly defined topic area so that the vocabulary that learners meet in the texts is limited because it relates to only one major topic” (Nation, 2013, p. 230).

There is some evidence showing that “sticking to one topic area results in a substantial reduction of vocabulary load” (Nation, 2013, p. 230). Vocabulary load refers to “the

vocabulary size necessary to understand individual texts”(Nation & Webb, 2008) Nation claims that the case for narrow reading is not overwhelming when it comes to “reducing vocabulary load and increasing repetitions”, however. He argues that Zipf’s law explains these “moderate vocabulary load and repetition benefits of narrow reading” because only a very small amount of vocabulary items accounts for a very large amount of the running words in any given text (2013, p. 231).

“Narrow reading may be more effective in increasing the amount of background knowledge that a learner brings to a text, thus aiding comprehension and as a result helping vocabulary learning” (Nation, 2013, p. 230). A recent study testing advanced L2 learner’s receptive and productive vocabulary learning after narrow reading, found that both receptive and productive vocabulary was acquired (Kang, 2015). “Repeated encounters with the thematic concept appeared to help learners develop semantic networks around the [target] words... Frequent encounters with target words in recurring contexts [also] helped their learning” (Kang, 2015, p. 175). It should be pointed out here that the vocabulary related to the theme of the topic was tested. As such, it can be expected that these words would also have a higher frequency across the texts the participants in this study read.

2.4.4.2 Vocabulary size and lexical coverage

Another important factor in relation to L2 vocabulary acquisition through reading is that of how many words learners need in order to comprehend the text they are reading. After much research in this field of study, most researchers will agree with Laufer (2010), who claims that learners should optimally know 98% of the vocabulary in the text they are reading; they should at least know 95% of the vocabulary in a text if comprehension is to occur.

In her 2010 study she also defines “the vocabulary threshold, that is the minimal vocabulary...necessary for ‘adequate’ reading comprehension” (Laufer, 2010, p. 15). Lexical coverage is defined by Laufer as “the percentage of words that a reader understands” (2010, p. 16). Again, most researchers agree that students seeking academic studies should have a knowledge of around 8,000 word families in order to comprehend the texts they must read at this level of study (98% lexical coverage). A vocabulary size of 4,000 – 5,000 word family is seen as a minimum, giving 95% lexical coverage (Laufer, 2010; Nation, 2006; D. Schmitt & Schmitt, 2012).

There is little research done directly in relation to vocabulary acquisition among Norwegian students (Langeland, 2012). Langeland conducted a longitudinal, qualitative study, over three years, in which students' vocabulary development was followed while they were between the ages of 9 and 13. The research questions related to tracking development of both receptive and productive vocabulary. Her study also implemented quantitative data gained through vocabulary tests and computer profiling programs in order to track student development. She found that for productive vocabulary use the students "[were greatly dependent upon] the first 1,000 words...but ... were gradually making use of a larger vocabulary" (2012, p. 140). Her study was conducted for a small group of students and there were not conclusive results, but she found that both receptive and productive vocabulary "developed unevenly and in spurts and plateaux, even regressions".

A relatively recent study (2008) of Danish students close in age to the target group of readers for the course materials analyzed in the current study may serve as an indication of vocabulary size relevant to first year high school students in Norway. Stæhr examined the vocabulary size of 88 fifteen and sixteen year old EFL learners. In the study, he examined possible correlations between vocabulary size as expressed in the improved 2001 version of the Vocabulary Levels Test (VLT) and the students' listening, reading and writing skills. The quantitative study found that the students "receptive vocabulary size was...strongly associated with their reading...abilities". He also found that most of the students did not have receptive, form-meaning knowledge of the first 2,000 frequency levels as defined by the GSL. Following the study, he postulates that "the 2000 vocabulary level is a crucial learning goal for low-level EFL learners" (Stæhr, 2008, p. 139). Stæhr excluded the academic word level from the VLT that was given to the students because "it is not relevant for low-level learners" (2008, p. 143). His comments are discussed further in section 5.4.2.

3. Methods and Materials

The aim of the current study is to analyze how general academic vocabulary is used in course textbooks written for advanced L2 learners of English in Norway and if these materials provide opportunities for the acquisition of this vocabulary. This thesis implements both computer programs and manual investigation of academic vocabulary use found in factual, textbook texts. The current study is based on the following research questions:

1. To what extent does the use of general academic vocabulary in factual, textbook texts provide the means for the implicit acquisition of this vocabulary during unassisted reading?
 - 1a. How is general academic vocabulary used within factual, textbook texts and across topic related texts?
 - 1b. To what extent does the use of glossaries in tailored texts assist advanced L2 English learners with the acquisition of general academic vocabulary?

This chapter provides an in-depth description of the research framework for the current study. Presented in this chapter are choice of research methods and research design, a more detailed presentation of methods used for data collection, a detailed description of the data analysis procedures, limitations to the study, and a discussion of validity and reliability.

3.1 Materials

The research materials consist of a survey and twenty-one factual texts obtained from three different textbooks commonly used in English subject classrooms throughout Norway (see section 3.4). There are seven texts analyzed from each textbook. Five of the texts are tailored and found in the textbook itself. Two texts from each book are authentic texts found on student websites connected to each textbook. Because the study is related to textbooks, there is an implied reader as participants and the texts themselves comprise the materials of investigation.

3.2 Methods

For any empirical study, choice of research method is largely dependent on the objectives of the study. As mentioned in the preceding chapter, Corpus Linguistics has been used in the current study as a method to find specific patterns in written input, and to provide the basis for

an investigation of questions concerning the broader discussion of vocabulary acquisition. The corpus examined here contains 21 textbook texts in three textbooks, providing a token basis of 28,734 for the materials analysis. Research methods applied to the current study have both quantitative and qualitative aspects. The distinction between these two research methods can be characterized as a continuum rather than two completely separate methods (Dörnyei, 2007). This is particularly true of the current study because the numeric data gathered here has been gathered using qualitative methods, i.e., the material is not from a representative population and at the same time, the statistics calculated from the numeric data have been used for in-depth descriptions of single cases. Quantitative methods have been employed to provide answers concerning AWL vocabulary usage patterns across the entire corpus. Qualitative methods have been applied to population choices and when examining questions linked to a more in-depth study of L2 vocabulary acquisition processes. As such, the research methods providing the methodological framework for the current study are a form of mixed methods (see section 3.2). The research questions guiding this thesis are both “specific, narrow, measurable, and observable” and at the same time “broad and general” (Creswell, 2014, pp. 27, 30). The investigation in focus examines specific patterns of vocabulary use in order to discuss this usage in terms of L2 vocabulary acquisition theory in general.

The research conducted for the current study can be classified as quantitative research methods for several reasons. The data collection process has been completed through the use of computer programs based on the GLS and AWL corpora, as well as the BNC and COCA corpora. From the data collected, the texts have been analyzed in order to investigate general patterns of AWL use. As such, the data collection process was quantitative in nature because the process was based on the use of instruments that provide a means to “analyze trends [and] relating variables...using statistical analysis, and [interpret] results by comparing them with ... past research, [while] taking an objective, unbiased approach” to the presentation of this numeric data (Creswell, 2014, p. 27).

Qualitative methods used in relation to the current study include the fact that the corpus used was a relatively small, purposeful sampling of the investigated population and should not be used to make general claims about all textbook texts for this target group. Purposeful sampling refers to the intentional selection of items to acquire information about the topic being researched (Creswell, 2014, p. 228). Also, in order to fully understand the relationship between variables, however, it has been necessary to limit the sample population. This has been done in order to provide the correct type of data for the numeric analysis. It has

also been necessary to choose the population from a set of specific criteria. For the current study, this means that the sampling was selected intentionally and not randomly (Creswell, 2014). The textbooks and texts analyzed for the current study were chosen from a set of criteria and not randomly selected (see section 3.4.1 and 3.4.2). It has been necessary to use a small population in order to “[develop] a detailed understanding” of the phenomenon in focus, find the correct type of information, and explain the relationships between variables in the relation to the research questions. These methods are related to qualitative research methods as defined by Creswell (2014, p. 228).

3.3 Research Design

A mixed methods research design has been used to “provide a better understanding of the research...question[s]...than either method [can accomplish] by itself” (Creswell, 2014, p. 565). The current study uses an embedded mixed methods design, with data that has been gathered sequentially i.e., in a certain sequence. The quantitative data gained through computer analyses is the main source of data for the current study. This data has been used to assess AWL vocabulary use; the qualitative methods applied through manual calculations and in-depth studies of single texts support the discussion of vocabulary usage and relate this to vocabulary acquisition. The supportive role that qualitative methods play in the study are what is meant by embedded mixed methods (Creswell, 2014, pp. 574-575). The following diagram outlines the data collection strategy used for the current study.

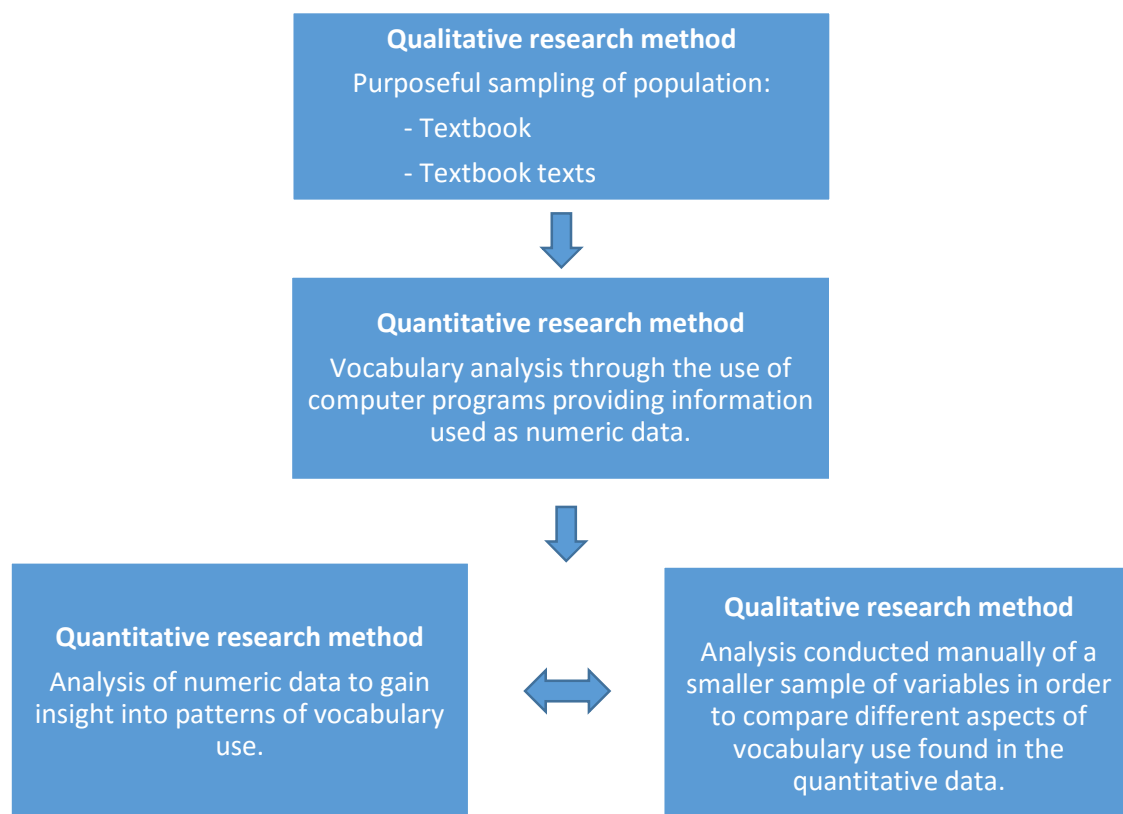


Figure 2. An embedded mix methods design, with data that has been gathered sequentially

3.4 Choice of Materials

For the current study, the target population is factual texts found in English course textbooks written for advanced L2 English learners in Norway. The population analyzed has been selected through purposeful sampling. Purposeful sampling is an example of qualitative data collection. This investigation contains three textbooks out of a total of nine written for this target group included in the analysis. Seven factual texts from each textbook were analyzed. As mentioned in section 3.2, the textbooks and texts have been chosen through a purposeful sampling of the population described more closely in the following. From these texts a small corpus of 28,734 words was developed and analyzed with the help of computer programs and manual calculations.

3.4.1 Textbook Choice

To determine which textbooks should be analyzed in the current study, a brief survey was conducted via e-mail involving 118 randomly selected schools throughout Norway's 19 municipalities. Representatives from the English department or library of each school was

asked to report what textbook or course materials, f. ex. NDLA (a digital gathering of texts) were used at their school. A total of 59 schools from 13 of Norway's municipalities, replied providing a response rate of 50 percent of schools replying in 68.5 % of the municipalities in Norway (see section 7.1). The information gathered here also corresponds to claims made by several publishers that their textbooks are some of the most widely used in Norway. The textbooks for this study were chosen from of the top three on the survey list. It was important to choose widely used textbooks in order for the information found in this study to be as relevant as possible to actual classroom settings. From the criteria explained above, the textbooks *Access to English* (Burgess & Sørhus, 2013), *Stunt* (Areklett et al., 2009) and *Targets* (Balsvik et al., 2015) have been used.

3.4.2 Text Choice

The text choices were also made using a purposeful sampling because general academic vocabulary is not used widely in works of fiction and discussions related to language learning topics vary widely across textbooks. (A. Coxhead, 2006; Nation, 2013). These texts have been chosen according to specific criteria. They had to be factual texts, examples of both tailored and authentic texts and topic related.

In order to meet the criteria for being categorized as a factual text, all of the texts had to be defined as factual texts in the index of the textbook. The tailor-made texts are texts written specifically for the textbook, often written by the textbook authors, but not always. They were found within the textbooks themselves. The authentic factual texts have been acquired through the website connected to each textbook.

The number of tailor-made factual texts included in the analysis is greater than the number of authentic factual texts in a ratio of five to two for each textbook. There are two reasons for this ratio. First of all, the number of authentic, written texts available on-line for each topic of focus were small. In most cases, there was only one written text available for each topic included in this study. Secondly, the tailor-made texts are more likely to be used by a majority of teachers since they are a central part of the textbook itself.

The texts chosen for analysis are directly related to two different competence aims outlined in the national curriculum for the English subject. These competence aims describe the students' needs to learn to "discuss and elaborate on the growth of English as a universal language" and to "discuss and elaborate on texts by and about indigenous peoples in English-

speaking countries” (Utdanningsdirektoratet, 2013). The competence aims were chosen because they represent two of seven subject aims related to the section for culture, society and literature (Utdanningsdirektoratet, 2013). As such, they are curriculum aims that are treated relatively extensively in the textbooks.

It was also important to have texts related to the same topics for each textbook in order to make a valid comparison of vocabulary use between textbooks and in order to examine general academic vocabulary use during narrow reading i.e., reading more than one text written about one specific topic. The investigation of narrow reading is directly related to implicit vocabulary acquisition and frequency of AWL word family occurrences across texts in other words, the range of vocabulary use.

3.5 Text Analysis

Each text has been analyzed with regard to the overall frequency and range of general academic vocabulary. The AWL word families included in glossaries have been examined for tailored texts only because the authentic texts have not used glossaries. The instrument used to gather the numeric data for the research includes three computer programs developed by Cobb and available on his website, the Lextutor (<http://www.lex Tutor.ca/>). These programs were chosen because their use is well-documented (Cobb, 2010), and they come highly recommended by other researchers in the field (Coxhead, 2012; N. Schmitt, 2010).

3.5.1 Token, Type and Word Family

As discussed in section 2.4.1, types, tokens and word families represent common ways of counting words in corpus studies. These word groups have been applied to the current study in different ways. To make the application of each word grouping for the current study as clear as possible, the ways in which they have been applied will be presented in relation to each separate text analysis in the following sections.

Tokens have been used when determining text length, such that text length is defined by how many tokens are used in a text. In the analysis of AWL vocabulary used in collocations each individual instance of word occurrence has also been used.

Word types have been used for the glossary analysis. There are two main reasons for my choice. By using word types, the exact word forms used in the glossary will be

represented to a greater extent, though this does also vary to some degree. Words such as *inhospitable* have a family head of *hospitable* so that the use of the word family distinction leads to a complete change in word meaning from that used in the context of the text. The word type distinction makes it easier to keep the analysis closer to word meanings within their proper contexts. The other reason for my choice involves counting the total number of glossary items. Each glossary term has been counted as one item unless the term has been defined in two different manners in the glossary, one example of this is *advert / advertisement* (Access to English, 2013, p. 67.). Here both terms have been counted separately since they are defined separately in the glossary. The individual word count has implications for the statistics related to grouping of vocabulary words used in the glossary (see Appendix 7.3). A more in-depth discussion related to this topic will follow in section 3.5.7.

In the text analyses for this study, word families have been used as a subdivision to determine frequency of AWL use in and across texts. My investigation of academic vocabulary use is related to advanced L2 learners and, as such, it can be expected that many derivatives of words will have a relatively low learning burden for these learners i.e., it should not take much effort to learn the derivatives. I am, however, aware that the use of word families can pose difficulties also for advanced L2 learners.

3.5.2 Text Preparation

There are several steps that have been taken to prepare the textbook texts for data analysis. Each text has gone through the same file conversion process, the files have then been treated for use in two corpus based, word analysis computer programs in which the material has been organized into data charts.

Each text had to be treated to allow for proper analysis before using the computer program. The texts have been scanned from the textbook to a pdf file. The pdf file has been converted into a word file, which has again been converted into a text only file. The text only file has then been treated for use in the VocabProfiler (VP) programs.

In accordance to the instructions provided on the Lextutor website, the text only files have to be changed to “[i]nclude an empty space after every comma or full stop any spelling errors occurring in the text must be corrected. Also it is important to adjust the text for the use of proper nouns” (Cobb, 2015). The software cannot recognize the different word classes so that tokens functioning as proper nouns may be counted as a content word (Nation &

Webb, 2011). Therefore, in the text analysis for this study proper nouns, such as names of specific people, places etc. (Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 288) have been manually formed into an explicit group.

By placing tokens used as proper nouns in a separate group in the analysis, the text coverage will be more accurately presented. Many proper nouns, such as *Australians*, will be familiar to ESL learners at this level of study and can therefore be recognized as vocabulary with a relatively low learning burden. By placing proper nouns in their own category, they are then “a factor in a text’s lexical density and [should be] factored into the calculation of text coverage” (Cobb, 2010, p. 187). For the current study, a list of each proper noun group is provided at the top of the text analysis and recorded in detail in each “Changes to Text” document provided for the text analysis of every text (see section 7.2). It is then possible to determine the level of word frequency needed to understand each text more without having a number of off-list words. It should also be noted that, for the current investigation, the use of proper nouns does not affect the count related to AWL vocabulary or the analysis of glossary terms. The role of proper nouns is therefore minimal for this study, but has still been considered so that the frequency level count may be discussed on like terms as other such analyses.

Proper nouns that occur in collocations have been treated in the following manner. Only the parts of the proper noun that would likely be recognized as off-list tokens were included in the list of proper nouns. The remaining tokens have been included in the original text for analysis. One example of such an occurrence is “the United States of America”, where only *America* is included on the proper nouns group list.

Other changes made for each text include spelling out contracted forms, omitting non-word forms such as pronunciation descriptions, removing hyphens in hyphenated words and separating compound words where the separate parts do not carry a difference in meaning to the whole. As a precaution, and to make the text analysis more reliable, all occurrences of contractions such as, *don’t*, *it’s* and *doesn’t* have been re-written as *did not*, *it is*, and *does not*. The decision to omit non-word forms related to pronunciation was made as a means of reducing the amount of “off-list” vocabulary for each text⁴. These forms are taken out only when they should be easily understood by the language learner. Examples from the texts

⁴ Off-list words refer to vocabulary that is not among the most frequently used BNC/COCA 25,000 frequency levels. By separating hyphenated words and compound words, the different word parts will be included on the frequency levels they belong to and will provide a more accurate evaluation of the vocabulary used in each text.

include groups of letters and spellings explaining pronunciation differences such as, “*tt*”, “*dd*”, and *Efrica*. The decision to remove hyphens is related to the inability of the computer program to process these vocabulary words. Many hyphenated words also occur as singular words and removing the hyphen tends not to have an effect on meaning (Nation & Webb, 2011). Examples of changes in hyphenated words are *working class*, and *day to day*. The separation of compound words has also been done to reduce the number of off-list words related to words that should be easily understood by the reader, but not processed by the computer program. This has only been done when the parts of these words do not change the meaning of the whole significantly, such as in *auto route*, and *loan words*. For each text there is a complete record of all changes made in the text appendixes (see section 7.2).

There are also some disadvantages to these text changes. The removal of hyphens does at times have an effect on the word count. The use of *non rhotic*, for example means that the prefix *non* stands on its own and is counted as a token. It seemed to be more important however, to let the root word *rhotic* play a part in the analysis rather than putting the entire word *non-rhotic* on the off-word list. One disadvantage of separating compound words came with the use of the word *lifestyle*. The separate parts do not change meaning in any great way so they have been separated as *life style*. *Style* is a part of the AWL, but *lifestyle* is not. No other compound words effected the AWL word family counts, however.

3.5.3 VocabProfiler (VP) Classic

Each text has been inserted into two different components of the Vocabprofil found on the Lextutor website. The first program, Vocabprofiler (VP) Classic, organizes lexical use in each text in the following manner:

It takes any text and divides its words into four categories by frequency:

(1) the most frequent 1000 words of English, (2) the second most frequent thousand words of English, i.e. 1001 to 2000, (3) the academic words of English (the AWL, 550 words that are frequent in academic texts across subjects), and (4) the remainder which are not found on the other lists. In other words, VP measures the proportions of low and high frequency vocabulary...in a written text, organizes the vocabulary in each text into word families, which are again arranged according to type, token and word frequency levels. (Cobb: <http://www.lextutor.ca/vp/research.html>)

The VP-Classic analyses are based on the 2,000 word families included in the GSL and Coxhead's AWL. These word lists were adapted into frequency levels by Laufer and Nation for use in computer software (Cobb, n.d.-b). For this study, the program has been used to examine AWL vocabulary usage in each text. The list of token, type and family frequencies obtained from the computer analysis has been compiled for each text (see section 7.2). This information has provided the numeric data necessary for statistics used in the discussion of L2 vocabulary acquisition presented in section 4.1.

The VP-Classic program was used to calculate the overall use of AWL word families as a percentage of the entire text. The calculations are done by dividing the total number of AWL tokens by the total number of tokens in each text. The calculations are provided in tables presented in the text analyses (see section 7.2). A comparison of average AWL word family coverage in each text was manually calculated in relation to each topic within the three textbooks, average AWL coverage for all of the texts in each textbook and an overall comparison between authentic and tailored texts across the textbooks (see section 4.1.1).

3.5.4 VocabProfiler (VP) Compleat

The VP-Compleat program has been used in order to have a complete frequency levels list related to each text. Here the AWL vocabulary is not specified, but the level of frequency for each token in the text has been determined using the BNC and the COCA corpora sorted into frequency levels (Nation, 2012). This analysis has been necessary in order to discuss the use of the cumulative percentage of lexical coverage in relation to reading comprehension.

The entire set of tokens within each text has been sorted into frequency lists and the percentages of use calculated by the VP-Compleat program (see section 7.2). Following this analysis the lexical coverage has been manually calculated for each text. These manual calculations were applied in order to place the total vocabulary use of each text into groups representing 95% and 98% lexical coverage. Manual calculations of average percentages have been conducted for use in comparisons between textbooks and between authentic and tailored texts (see section 4.1.2).

3.5.5 Combining Both Programs

In order to provide more in-depth information concerning AWL vocabulary usage in the corpus materials, it was important to analyze this vocabulary with both the VP-Classic and VP-Compleat programs. The VP-Compleat analyses categorized AWL word families found in the corpus into BNC/COCA frequency levels, which helped determine to what extent advanced L2 learners might be expected to have prior knowledge of these terms.

The statistics from this list were used to conduct manually calculations of average frequency levels for the AWL vocabulary in each text. The accumulated information was used in comparisons of textbooks and between authentic and tailored texts (see section 4.1.3). It was used for an analysis of AWL word families only occurring once in each text (see section 4.1.4) and across topic related texts (see section 4.1.4.2). The manual calculations have been executed in order to provide more precise information in relation to how often it might be expected that particular AWL word families occur in a text and what level of learning burden may be expected for these word families.

3.5.6 Range Analyses

For the current study, effects of vocabulary usage in texts with similar topics have been analyzed, making it possible to discuss these texts in relation to research related to narrow reading. Three to four texts covering two different topics have been chosen from each of the three textbooks in the study. The occurrences of AWL vocabulary across all texts related to a topic within a textbook was studied through the use of the Range program found on Cobbs' website (lextutor.ca/). In this program the word only files (see section 7.4) related to specific topics. The analysis for each textbook contained three or four different texts for each topic. The two topics chosen were related to two English subject competence aims i.e., English as a global language and indigenous peoples. In the Range program all of the word tokens in each set of texts are sorted into word families into groups pertaining to frequency of occurrences across the texts, BNC frequency levels and the number of texts the word family occurs in.

The choice to use word families and not word types for this investigation has been taken to give a more correct representation of the number of repetitions of tokens that should easily be understood by L2 learners. By doing so such tokens such as *area/areas* are grouped together and counted in one frequency group.

Information from the Range analyses was manually sorted to only include AWL vocabulary for each topic and each textbook (see section 7.4). The material was then organized into charts that show statistics related to the amount of ALW vocabulary that is found in several texts, the frequency of the AWL vocabulary across the texts for each topic and an in-depth discussion of the AWL word family range. These statistics were calculated manually (see section 4.1.4.2).

3.5.7 Glossary Analyses

The three textbooks included in the current study had similar ways of organizing the glossaries. In the textbook *Access to English* glossaries are found in the margin of the text. They are written such that the English words are in italics, followed by a Norwegian translation that is not italicized (Burgess & Sørhus, 2013). Glossaries are also sometimes organized in the margins of *Stunt*, but are also found at the end of some of the texts. The English words are written in bold type, followed by Norwegian translations that are italicized (Areklett et al., 2009). *Targets* also places the glossary terms in the margins of each text. The glossaries are written such that the English words are in bold type, followed by Norwegian translations. This textbook also includes some English definitions of terms in the margins, none of which occur on the AWL (Balsvik et al., 2015). The authentic texts did not include the use of glossaries and are therefore excluded from this section of the study. In the analysis, vocabulary used in each text glossary has been categorized according to frequency levels and the AWL.

For each glossary used in the tailored texts three analyses have been conducted. The first analysis determined glossary coverage, the second analysis determined the BNC/COCA frequency levels of all the analyzed texts and the final analyses exemplified AWL vocabulary usage in the glossaries. Overall glossary coverage was manually calculated by dividing the total number of tokens in each text by the total number of glossary tokens used in each text (see sections 7.2 and 7.3). These analyses are based on research findings and recommendations from linguists as to how glossary items may best help L2 learners expand their comprehension of a text by helping them towards greater understanding of vocabulary.

If a glossary word type was used more than once in the glossary, such as *eventually* (see section 7.2.1.1.3), all tokens were counted in the calculation. If the word type used in the glossary occurred also in inflected forms of the headword, all of the tokens from the entire

word family were counted, such as *spell* and *spelling* (see section 7.2.2.2.3). Each token was counted also for closely related derived forms like *communicate* and *communication* (see section 7.2.2.1.2). Collocations have been counted as separate words here, but only as one glossary item in the analysis of glossary use compared to frequency levels. An example is the collocation *with regard to* that has been counted as three tokens to determine the overall percentage of glossary use in each text. I have done so to keep with the word division of token in both sections. Since text length is defined in tokens per text, I felt it was most accurate to also count the glossary items in terms of tokens used. However, collocations have been counted as one glossary item in the analysis of glossary items describing use of collocations. An in-depth presentation of these findings is provided in section 4.2.1.

After calculating the glossary coverage rates, all of the glossary items for each text were analyzed with the VP-Compleat program. This analysis was used to provide information concerning the BNC and COCA frequency levels of the glossary terms in order to examine if the glossary terms may be expected to be beyond the vocabulary size of the readers. The numeric information from these analyses are presented in detail per text and per textbook in section 4.2.2.

The final glossary analysis focuses on the use of AWL word families in the glossaries. Several different analyses were conducted in relation to these words. First, an analysis has been conducted with the help of the VP-Classic program in order to determine which glossary items are on the AWL. The VP-Compleat has then been used to determine what frequency levels these AWL words occur at. These investigations have been conducted in order to examine to what extent AWL vocabulary occurs in the glossaries and in order to determine more precisely what types of AWL vocabulary are found in the glossaries. These variables are directly related to L2 vocabulary acquisition of general academic vocabulary (see section 2.4).

An investigation of AWL vocabulary items occurring once, both with relation to AWL glossary items occurring only once and overall AWL vocabulary occurring only once in the entire text has also been conducted with the use of the computer programs. Manual calculations using information from the program data have been conducted for all of the mentioned computer analyses in order to determine percentages of occurrences related to the different variables. These investigations have been completed to examine the extent of AWL vocabulary usage in order to determine if this promotes implicit vocabulary acquisition. A detailed presentation of these analyses and the findings there of are found in section 4.2.2.

3.6 Ethical Issues

Ethical issues in conducting research involve showing respect for the audience and respecting the participants. These issues are important to consider during the entire research process (Creswell, 2014). There are still ethical considerations to be made when using textbooks and not individuals as participants. The work of others during the creation of these course materials needs to be treated with respect. This means that any analysis conducted using these materials needs to be treated fairly. Any data collected also needs to be “reported honestly, without changing or altering the findings to satisfy certain predictions” (Creswell, 2014, p. 38). Results should be reported in a balanced manner and any reference to work conducted by others must be cited properly (Creswell, 2014).

Ethical issues have been taken into consideration both during the data collection process and in the discussions related to the findings in order to be aware of the effects the present research may have on others (Cohen, Manion, & Morrison, 2007). I have tried to keep the data collection process as transparent as possible so that others may also repeat the investigations that have been made. This is the reason that the appendices section is so comprehensive. I have done my best to cross check all of the analyses to make sure that my findings reflect an honest report of the information collected. I have also tried to provide a balanced discussion of the findings to show respect for those that have produced these materials and the work they have done in their creation.

The project does not treat any form of personal data and is therefore not subject to notification requirements provided by the Norwegian Social Science Data Services (NSD).

3.7 Reliability and Validity

Reliability and validity are paramount aspects of any research. These are two concepts that “are bound together in complex ways”. If the study is not reliable, it is not valid (Creswell, 2014, p. 177). Due to the overriding use of numeric data in the current study these two terms will be used as readily defined in quantitative methods, without the need for terms such as trustworthiness or credibility (Cohen et al., 2007; Dörnyei, 2007). Reliability in connection with collecting quantitative data refers to the stability and consistency of data provided from the use of the data collection instrument applied (Creswell, 2014). Validity on a very basic level relates to if instruments used in the research process actually measure what they set out to. Validity can also be related to use of triangulation to improve the “richness and scope of

the data” as well as appropriate sampling and appropriate treatment of the data (Cohen et al., 2007, p. 133). For the current study, it has been necessary to analyze those texts in textbooks that would compare to authentic genre in which AWL vocabulary could be expected to a greater degree than in works of fiction. This is why I have chosen only to analyze factual texts. I have also used triangulation to be able to provide more in-depth data that shows how general academic vocabulary is use in a way that enhances implicit vocabulary acquisition. Three different computer programs have been applied to the corpus to examine different traits in vocabulary use. I have also analyzed glossaries as even one more way of discovering important information regarding how textbook texts may help learners acquire AWL vocabulary implicitly.

In the current study, reliability is closely related to the choice of computer programs and the use of these programs to provide reliable data. The data collection instrument used for the current study i.e., Lextutor, was chosen because its use has been well-documented from previous research. Prominent researchers continue to recommend the use of these computer programs (Coxhead, 2012; Nation, 2013; N. Schmitt, 2010). The programs have repeatedly produced the same results when being tested for this study as well.

Some of the methods used to treat the texts before analysis can lead to challenges in how words are counted. In vocabulary assessment studies there are few standard measurements of validity. Therefore, the validity of this research will often times have to be demonstrated in the way the study preforms on its own. One step in the process is to specify the content i.e., content validity (N. Schmitt, 2010). The current study has placed focus on the study of general academic vocabulary, at term that has been carefully described and discussed in relation to relevance to the research population and in relation to operationalization (see section 2.1). All changes to the texts have been recorded so as not to weaken the validity of the study (N. Schmitt, 2010). As discussed in section 3.5.2, it has been necessary to remove hyphens in hyphenated words and separate compound words so that they would not be placed in the off-list category. Too many word families on the off-list category would give an inaccurate account of overall frequency level patterns, but has little relevance to the AWL analyses. Many hyphenated words commonly occur as singular words and removing the hyphen tends not to have an effect on meaning (Nation & Webb, 2011). Examples of changes in hyphenated words in the current study are *working class*, and *day to day*. The separation of compound words has done in relation to words that should be easily understood by the reader, but not processed by the computer program. Again, it has been important to record all changes

made to each texts so that others may see what has been done (see Appendix 7.2). One separation of compound word has affected the AWL analysis and that is the word *lifestyle*. A more in-depth discussion of this vocabulary word and how it has been dealt with for the current study is found in sections 3.5.1 and 4.1.3.2. Changes made to the text only files have been done manually and can therefore contain mistakes. I have, during the data collection process, documented all text changes and have also double checked the text only files, but with 21 texts analyzed in several ways, mistakes may have occurred without being detected.

When first using the VP-Compleat program, I was unable to gain the correct data with regard to the analysis of proper nouns. The program designer, Cobb, was kind enough to help determine the cause of my difficulties and even changed it to accommodate for the use of commas in the proper nouns section. The reliability of the text analyses using the Range and VP programs is therefore high.

The validity of the current research is dependent on reliable data, something I am confident I have done as much as possible to secure for my research. For this study, validity is related to two aspects of the investigation; both how AWL vocabulary was used and if the use of this vocabulary supports L2 vocabulary acquisition of AWL word families through unassisted reading. First, it is important to establish if the instruments used correctly measure AWL vocabulary usage. Triangulation has been used to better describe overall AWL vocabulary usage and thus enhance the validity of the results for the current study (Dörnyei, 2007, p. 165). This has been done by analyzing frequency of occurrence in and across texts, the frequency levels of AWL vocabulary and AWL vocabulary use in glossaries. Triangulation refers to the “process of corroborating evidence from different ...types of data [and] methods of data collection...in descriptions and themes in qualitative research” (Creswell, 2014, p. 13). Secondly, the investigation of AWL vocabulary acquisition has been dependent on applying vocabulary acquisition theory to the concrete data that was accumulated. When applying theory and hypotheses to a discussion there are always some assumptions that must be considered. In the current study, it is assumed that six repetitions of an AWL word family within or across texts will help L2 learners acquire these terms. It is also assumed through empirical evidence (Nation, 2013; D. Schmitt & Schmitt, 2012), that word families occurring at higher frequency levels may be less well known to advanced L2 learners. Another assumption is that lexical coverage is an important factor in relation to reading comprehension. These assumptions and their importance for the current study have been discussed in more detail throughout chapter two. By relating the discussion of L2

vocabulary acquisition directly to theory and hypotheses and discussing how these have been investigated in previous research, I hope that the validity of my research is strong.

3.8 Limitations

There are several limitations to the current study. In terms of the materials used, the study is limited in focus to one, small part of the vocabulary acquisition process i.e., the assessment of vocabulary usage in written English course materials. Vocabulary focus is also limited to include only general academic vocabulary as represented in the AWL. The research questions have limited the investigation to AWL vocabulary usage and ways in which written input may provide the means for implicit vocabulary acquisition of these words. In relation to the theoretical framework, the study has been limited to usage-based methods and hypotheses relevant to implicit vocabulary acquisition during unassisted reading. The discussion of word knowledge has been limited to form-meaning, receptive knowledge because the analyses investigate content in written input and not L2 learner abilities.

4. Results

The following chapter will provide an overview of the findings for this study, a brief explanation of the possible importance of these findings, and a presentation of evidence in support of the results. Focus is placed on providing answers to the main research questions guiding the thesis. The two underlying questions related to AWL use in general and AWL use in glossaries will be addressed first. Afterwards, the main research question related to implicit vocabulary acquisition will bring the analyses for the current study together.

4.1 AWL Vocabulary Use

The first underlying research question posed was how AWL vocabulary has been used in the factual, textbook texts making up the 21 factual texts, containing a total of 28,734 tokens in the corpus of the current study. Several different analyses have been conducted in order to provide data related to this part of the research. In the present section, findings will be presented in relation to three factors dealing with overall vocabulary use in the analyzed texts. The first findings presented are concerning AWL word family coverage i.e., how many AWL tokens are used in comparison to the total number of tokens in each text. The second set of finds reported on here deal with a presentation of AWL word families with a frequency of six

times or more repetitions. These findings will be presented for both in-text frequency and through a Range analysis across topic related texts. Finally, the third factor applying to general academic vocabulary usage presents the findings for the AWL word families that have only occurred once in and across texts.

4.1.1 Percentage of Total Text

In the following analysis, the total percentage of AWL vocabulary used in the analyzed texts has been investigated and compared to what may be expected in other non-fictional genre. Authentic newspapers normally provide around 5% coverage of AWL vocabulary, while most academic texts written for university studies provide between 8% -10% coverage (Coxhead, 2006; Coxhead, 2000; Nation, 2013). This has been done in order to investigate how closely the general academic vocabulary in course materials resembles academic texts students may encounter later in their studies.

There was a combined total of 1521 AWL tokens used in the corpus of 28,734 tokens, providing a total average of 5.3% AWL coverage (see tables 1 and 2)⁵. This shows that, on average, AWL vocabulary was used at the same rate often found in English language newspapers. Rather large variations between texts, topic subjects and textbooks were also found. Differences in percentage of AWL vocabulary used varied from 2.3% to 8.4%. There was greater use of AWL vocabulary in authentic texts compared to tailored texts. On average, AWL vocabulary use in the tailored texts was relatively low, at 4.1% but still just under what can be found in English language newspapers. The authentic texts used an average of 7.5% AWL vocabulary, slightly below what could be expected in academic texts.

Table 1. *Total corpus average of AWL used per text for tailored texts, in percentage*

Tailored texts	Total tokens	AWL tokens	AWL %
Divided by a Common Language	1064	24	2.3
A Global Language	2075	48	2.3
Aboriginal Australians	1704	69	4.1
Native Americans: Original Inhabitants	1365	49	3.6
Stolen Children	297	10	3.4
Australia – the Island Continent	1506	105	7.0
Native Americans: We Are Still Here	1410	79	5.6

⁵ All of the data for the calculations discussed in relation to overall AWL in-text use have been provided by the AP-Classic analyses and are found in Appendices 7.2. The averages have been calculated manually.

British vs. American English	1104	26	2.4
English as a World Language	566	22	3.9
The Power of English Part 2	1866	97	5.2
Native Americans	900	32	3.6
Australia: The Birth of a Nation	848	22	2.6
Stolen Generation	461	26	5.6
The Flavours of English	2185	74	3.4
The Power of English Part 1	1161	75	6.5
Mean score:	18512	758	4.1

Table 2. Total corpus average of AWL used per text for authentic texts, in percentage

Authentic texts	Total tokens	AWL tokens	AWL %
Effects of Removal	5595	412	7.4
English and the Future	896	63	7.0
There Is an Epidemic	868	71	8.2
Native Americans In Business	621	52	8.4
“Indian” Mascots	1502	122	8.1
Renaming English	740	43	5.8
Mean score:	10222	763	7.5

Due to large variations, each textbook will be discussed and compared in the following.

4.1.1.1 Access to English

Two out of seven texts analyzed in *Access to English* (28.6%) used over 5% AWL vocabulary. Variations in the remaining texts fall between 2.3% and 3.6% AWL vocabulary coverage. Clear differences between AWL vocabulary used in tailored and authentic texts were found. Average AWL vocabulary coverage in the tailored texts was relatively low, at 3.1%. The authentic texts averaged, 7% which is slightly below vocabulary expectations in academic texts. There was a combined total of 295 AWL tokens of 7866 tokens used in these textbook texts, giving a total average of 3.1% AWL coverage for *Access to English*. This was the lowest average rate of the three textbooks analyzed in the current study.

Table 3. *Access to English*: Total average of AWL per tailored text, in percentage

Tailored texts	Total tokens	AWL tokens	AWL %
Divided by a Common Language	1064	24	2.3

A Global Language	2075	48	2.3
Stolen Children	297	10	3.4
Native Americans: Original Inhabitants	1365	49	3.6
Aboriginal Australians	1704	69	4.1
Mean score:	6505	200	3.1

Table 4. *Access to English: Total average of AWL per authentic text, in percentage*

Authentic texts	Total tokens	AWL tokens	AWL %
Renaming English	740	43	5.8
Native Americans In Business	621	52	8.4
Mean score:	1361	95	7.0

4.1.1.2 *Stunt*

Three out of seven texts (42.9%) contained over 5% AWL coverage. The tailored texts used an average of 3.3% AWL vocabulary, slightly lower than that used in *Access to English* but still under rates used in English language newspapers. The authentic texts showed 7.5 % AWL coverage and, as such, are close to what would be expected in academic texts and only slightly below rates found in *Targets*. There was a combined total of 611 AWL tokens in the corpus of 10342 tokens, providing a total AWL coverage for *Stunt* of 5.9 %, highest of all three textbooks, but only very slightly higher than *Targets* at 5.8%.

Table 5. *Stunt: Total average of AWL per tailored text given in percentage*

Tailored texts	Total tokens	AWL tokens	AWL %
British vs. American English	1104	26	2.4
English as a World Language	566	22	3.9
Native Americans	900	32	3.6
Australia: The Birth of a Nation	848	22	2.6
Stolen Generation	461	26	5.6
Mean score:	3879	128	3.3

Table 6. *Stunt: Total average of AWL per authentic texts given in percentage*

Authentic texts	Total tokens	AWL tokens	AWL %
There Is an Epidemic	868	71	8.2
Effects of Removal	5595	412	7.4
Mean score:	6463	483	7.5

4.1.1.3 Targets

Six out of seven texts (85.7%) provide over 5% AWL coverage. The tailored texts use, on average, 5.3% AWL vocabulary, by for the highest average of the three textbooks used for this study. The two authentic texts have an average of 7.7% AWL coverage; also the highest rate of all three textbooks. The AWL coverage for authentic texts are close to what would be expected in academic texts. The combined average for Targets at 5.8% (615/10526) use of AWL vocabulary is slightly lower than the combined average for *Stunt*. The rates of AWL coverage are more evenly divided among the texts in *Targets*. In *Stunt* the authentic text that was over 5000 tokens (see table 6) has colored the results for this textbook.

Table 7. *Targets: Total average of AWL per tailored texts given in percentage*

Tailored texts	Total tokens	AWL tokens	AWL %
The Flavours of English	2185	74	3.4
The Power of English Part 1	1161	75	6.5
The Power of English Part 2	1866	97	5.2
Native Americans: We Are Still Here	1410	79	5.6
Australia – the Island Continent	1506	105	7.0
Mean score:	8128	430	5.3

Table 8. *Targets: Total average of AWL per authentic texts given in percentage*

Authentic texts	Total tokens	AWL tokens	AWL %
English and the Future	896	63	7.0
“Indian” Mascots	1502	122	8.1
Mean score:	2398	185	7.7

Exposure to general academic vocabulary is essential during unassisted reading, in order to facilitate implicit L2 vocabulary acquisition. All of the texts in the current study would expose advanced L2 learners to AWL vocabulary, authentic texts more so than tailored texts. The analyses presented above show that there are large differences between texts, but the overall differences between textbooks is not as great, though *Targets* has a far greater number of texts with AWL coverage levels around 5%. As expressed earlier, the L2 vocabulary process is too complex to assume that a high percentage of AWL use in itself will provide the means for implicit vocabulary acquisition during unassisted reading. The next analyses presented examined more closely how the AWL vocabulary found in the corpus was used.

4.1.2 In-text Frequency

While the first analyses determined to what extent AWL vocabulary was used within the corpus, the second set of analyses investigate more closely how repletion of AWL word families was used in each text. Taking into consideration previous research used to establish theory related to the possibility of implicit vocabulary acquisition through unassisted reading, the following analyses have been conducted to examine the extent to which AWL word families were repeated six times or more in one text. The use of this baseline number relates to previous recommendations from research regarding implicit learning through unassisted reading (see section 2.4.2).

Table 9. *AWL word family in-text frequency of six or more repetitions, in percentage.*

Textbook texts	AWL % of total*	AWL total word fam.	word families 6+	6 + %
Textbook: Access to English				
Tailored text: Divided by a Common Lang.	2.3	18	0	0.0
Tailored text: A Global Language	2.3	34	0	0.0
Authentic text: Renaming English	5.8	29	0	0.0
Tailored text: Native Americans: Original In.	3.6	36	0	0.0
Tailored text: Aboriginal Australians	7.0	51	1	2.0
Tailored text: Stolen Children	3.4	8	0	0.0
Authentic text: Native Americans In Business	8.4	26	2	7.7
Textbook: Stunt				
Tailored text: British vs. American English	2.4	17	0	0.0

Tailored text: English as a World Language	3.9	18	0	0.0
Authentic text: There Is an Epidemic	8.2	47	0	0.0
Tailored text: Native Americans	3.6	23	1	4.3
Tailored text: Australia: The Birth of a Nation	2.6	17	0	0.0
Tailored text: Stolen Generation	5.6	17	0	0.0
Authentic text: Effects of Removal	7.4	157	9	5.7
Textbook: Targets				
Tailored text: The Flavours of English	3.4	51	1	2.0
Tailored text: The Power of English Part 1	6.5	44	2	4.5
Tailored text: The Power of English Part 2	5.2	51	3	5.9
Authentic text: English and the Future	7.0	48	0	0.0
Tailored text: Native Americans: Still Here	5.6	58	0	0.0
Tailored text: Australia – the Island Continent	7.0	64	1	1.6
Authentic text: “Indian” mascots	8.1	67	3	4.5
Mean score:		881	23	2.6

*AWL percentage of total is simply a reminder of the total AWL in-text coverage.

An examination of the entire corpus showed that, on average, 2.6% of the AWL vocabulary in this corpus occurred six times or more. Twelve of the 21 analyzed texts (57%) did not contain AWL word families recycled the minimum of six times. In two of the textbooks used for the current study, five of the seven analyzed texts (71.4%) did not repeat any AWL word families six times.

There were a total of 23 AWL word families used six or more times in the separate textbook texts. Four of the same word families occurred six times or more in two or all three of the textbook texts analyzed in this study. These are *area*, *economy*, *remove*, and *culture*. In relation to frequency levels, *area* is categorized in the 1,000 frequency level, and *economy*, *remove*, and *culture* are categorized in the 2,000 frequency level. Of the AWL headwords repeated six or more times, only *immigrate*, *globe*, *communicate* and *media* are found outside of the 2,000 frequency level. A VP-Compleat analysis of the recycled AWL word families revealed that a majority of the word families (56,5%) were found at or below the 2,000 frequency level. BNC and COCA frequency levels can inform this research in terms of placing these word families within the investigation of lexical coverage. A further discussion of this aspect of the study will be pursued in section 4.3.

The remaining discussion of this analysis has been organized through a presentation of each recycled AWL word family represented in the corpus of 21 texts.

Table 10. *Access to English: List of AWL headwords and word types used six or more times*

Texts	Headword	Word types
Text 1	<i>area</i>	<i>area</i> (3), <i>areas</i> (4)
Text 2	<i>economy</i> , <i>community</i>	<i>economic</i> (6) <i>community</i> (4), <i>communities</i> (5)

In the textbook *Access to English* there were two of the seven texts (28.6%) that had six or more repetitions of AWL vocabulary. A majority of the texts (71.4%) analyzed in the current study, for this textbook, did not have any AWL word families that were recycled six or more times. In the tailored text (text 1) “Aboriginal Australians”, the head word *area* was repeated seven times during the text; three times as *area* and four times as *areas*. In the authentic text (text 2) “Native Americans in Business”, the two word families, *economy* and *community* were repeated six and nine times respectfully. Community was also used in singular and plural forms, while the word type *economic* was repeated all six times in this text (Burgess & Sørhus, 2013). A total of three word families from the AWL that were repeated six or more times in the seven texts analyzed for this textbook. The overall in-text repetition of AWL word families in *Access to English* is the lowest of the three textbooks included in this study.

Table 11. *Stunt: List of AWL headwords and word types used six or more times*

Texts	Headword	Word types
Text 3	<i>remove</i>	<i>removal</i> (7)
Text 4	<i>adapt</i>	<i>adapt</i> (6), <i>adapted</i> (1), <i>adaptation</i> (1)
	<i>area</i>	<i>area</i> (5), <i>areas</i> (3)
	<i>create</i>	<i>created</i> (5), <i>creation</i> (1)
	<i>economy</i>	<i>economy</i> (3), <i>economic</i> (5)
	<i>environment</i>	<i>environment</i> (16), <i>environmental</i> (1), <i>environments</i> (10)
	<i>policy</i>	<i>policy</i> (8)
	<i>region</i>	<i>region</i> (3), <i>regions</i> (6)
	<i>remove</i>	<i>remove</i> (1), <i>removal</i> (40), <i>removals</i> (1), <i>removed</i> (2)
	<i>tradition</i>	<i>tradition</i> (3), <i>traditions</i> (3), <i>traditional</i> (5)

The textbook *Stunt* also had two texts (28.6%) in the analysis that contained AWL vocabulary repeated six or more times. A majority (71.4%) of the texts analyzed in *Stunt* did not have any AWL word families that were recycled six or more times. In the tailor made text (text 3) “Native Americans” the word family *remove* was repeated a total of seven times; each

repetition was represented by the word type *removal* (Areklett et al., 2009). The authentic text (text 4) “Effects of Removal on American Indian Tribes” was a long text, 5595 tokens, that had a total of nine word families that were repeated six or more times. The least amount of repetitions was eight, and the greatest amount of repetitions was 44 (*remove*). The following word families are represented in the text: *adapt*, *area*, *create*, *economy*, *environment*, *policy*, *region*, *remove*, and *tradition*. The word family *adapt* was represented by *adapt* (6), *adapted* (1) and *adaptation* (1). *Area* was represented in singular (5) and plural (3) forms. The headword *create* was represented as *created* (5) and *creation* (1). Economy The word types *economy* (3) and *economic* (5), *environment* (16), *environmental* (1) and *environments* (10) were also used in the text. The headword *policy* was used as *policy* eight times in the text. The headwords *region*, *remove*, and *tradition* have been used in the following manner: *region*, singular (3) and plural forms (6); *remove* (1), *removal* (40), *removals* (1), *removed* (2); *tradition* (3), *traditions* (3) and *traditional* (5).

Table 12. *Targets: List of AWL headwords and word types used six or more times*

Texts	Headword	Word types
Text 5	<i>culture</i>	<i>cultural</i> (1), <i>culture</i> (3), <i>cultures</i> (2)
Text 6	<i>communicate</i> <i>globe</i>	<i>communicate</i> (1), <i>communication</i> (7), <i>communications</i> (1) <i>global</i> (8), <i>globe</i> (1)
Text 7	<i>culture</i> <i>economy</i> <i>immigrate</i>	<i>cultural</i> (2), <i>culturally</i> (1) <i>culture</i> (2) <i>cultures</i> (1) <i>economy</i> (2), <i>economic</i> (4), <i>economically</i> (1) <i>immigrant</i> (1), <i>immigrants</i> (5) <i>immigration</i> (3)
Text 8	<i>percent</i>	<i>percent</i> (8)
Text 9	<i>culture</i> <i>team</i> <i>media</i>	<i>cultural</i> (2), <i>culture</i> (8), <i>cultures</i> (1) <i>team</i> (18), <i>teams</i> (2) <i>media</i> (6)

Five out of seven texts (71.4%) analyzed for *Targets* contained AWL word families that were repeated six or more times. In the tailor made text (text 5), “The Flavours of English” the word family *culture* is used six times in the following way: *cultural* (1), the singular (3), plural (2) forms of *culture*. Other tailor made texts in this category include “The Power of English” parts 1 and 2 (texts 6 and 7) and (text 8) “Australia – the Island Continent” (Balsvik et al., 2015). In the two “Power of English” texts, a total of five headwords were use six or more times. Word families represented in these two texts were *communicate*, *globe*, *culture*, *economy* and *immigrate*. The following word types were used: *communicate* (1), *communication* (7), *communications* (1); *global* (8), *globe* (1); *cultural* (2) *culturally* (1) *culture* (2) *cultures* (1); *economy* (2), *economic* (4) *economically* (1). The text about Australia

included one headword *percent*, which was also used as *percent* in the text eight times. In the authentic text (text 9) “Indian Mascots” the word families *culture*, *team* and *media* have been used six times or more in the following ways: *cultural* (2), *culture* (8), *cultures* (1); *team* (18), *teams* (2); *media* (6). *Targets* was the only textbook that had a majority of texts in which some AWL word families were recycled six or more times.

L2 learners may also be assisted in implicit vocabulary acquisition by reading several texts concerning the same topic, narrow reading. For the next set of analyses, the investigation has been centered on determining if the number of AWL word families recycled six or more times will be greater across three and four topic related texts.

4.1.3 Range

Findings related to the text analyses presented in this section of the thesis examine to what extent narrow reading can enhance implicit vocabulary acquisition of AWL word families. The analyses have been conducted using the Range computer program found on lextutor.ca (see section 3.5.6). Topic related texts have not been analyzed across textbooks, since it is unlikely that more than one textbook would be used in a classroom situation.

Table 13. *AWL word families occurring across topic related texts for total corpus*

Textbook and Topics	AWL total word fam.	word families 6+	6 + %
<i>Access to English</i>			
Global English	81	2	2.5
Indigenous peoples	121	4	3.3
<i>Stunt</i>			
Global English	82	1	1.2
Indigenous peoples	214	17	7.9
<i>Targets</i>			
Global English	194	8	4.1
Indigenous peoples	189	7	3.7
Average for total corpus	881	39	4.4

As could be expected, a Range comparison of texts across each topic shows that there was a slight increase in the number of AWL word families recycled six or more times. All of the topics had at least one AWL word family that occurred six or more times. The percentage of word families in the Range analysis that occurred six times or more rose from 2.6% per

text (see table 9) to 4.4%. However, these rates are still very low, with a high of 7.9% and a low of 1.2% per topic. A description of AWL vocabulary for each textbook provides a more detailed understanding of the effect narrow reading may have on frequency of AWL vocabulary and will be presented in the following. The data described here can be found in Appendix 7.4.

4.1.3.1 Access to English

In relation to the topic of global English, the two word families *globe* and *culture*, occurred six or more times, six and eleven times to be exact. The previous analysis of in-text frequency had found no AWL word families used six times or more.

For the topic of indigenous peoples, there were a total of four word families with a frequency of six or greater. These four word families were: *community*, *area*, *economy* and *conflict*. Compared to the in-text frequency analysis the AWL word family *conflict* has also reached a frequency of six occurrences in the Range analysis. The other word families have increased their frequency of occurrences to eleven, nine, and seven respectively.

4.1.3.2 Stunt

For the topic of global English, one word family, *communicate*, was repeated six times. The in-text frequency analysis for this topic did not contain AWL word families repeated six or more times.

Before discussing the topic of indigenous peoples for this text it is important to point out that the authentic text related to the topic had a far greater amount of AWL word family frequency than any other text in the analysis, with 412 tokens of AWL words used. It was also, by far, the longest text analyzed (5595 tokens) which most likely accounts for the differences in frequency here. The largest number of AWL word family repetitions was 53 (*remove*), while six other AWL word families occurred between 10 and 27 times. The group of 17 AWL words families for this category were: *remove*, *environment*, *area*, *adapt*, *tradition*, *establish*, *region*, *policy*, *create*, *identify*, *economy*, *culture*, *style*, *community*, *define*, *final*, *resource*. The number of AWL word families recycled six times or more increased from nine to seventeen compared to the in-text frequency analysis. The additional word families were: *establish*, *identify*, *culture*, *style*, *community*, *define*, *final*, *resource*.

The use of *style* here represents a weakness with the current study because the type *lifestyle* has been divided in the text analysis in order to avoid a large number of off-list words. While *style* is a head word on the AWL, *lifestyle*, is not a part of the word family defined by Coxhead (n.d.). The result has been that one AWL word extra has been counted in the total usage for this category. As such, the statistic is slightly higher than it should be. Since this is only the case with one compound word in the analyses, I have not adjusted the statistics.

4.1.3.3 Targets

For the topic of global English, as presented in *Targets*, nine AWL word families had a frequency of six or more repetitions: *communicate*, *culture*, *economy*, *region*, *technology*, *establish*, *area*, *immigrate* and *globe*. Compared to the in-text frequency analysis, the frequency of three word families had risen to over ten repetitions; these were, *culture*, *communicate* and *globe*. The highest frequency rate for the previous in-text analysis was nine.

The topic related to indigenous peoples had the greatest number of AWL word families, seven occurring across all three texts. Also 3.7% of the AWL word families had a frequency of six or more repetitions; three of these being repeated ten or more times. These AWL headwords were: *team*, *culture*, *percent*, *area*, *federal*, *challenge*, *media*. The AWL word families *team*, *culture*, and *percent* were repeated 21, 14 and 10 times respectively. In relation to the in-text frequency analysis for this topic, three new AWL word classes have a frequency of six or more: *area*, *federal*, *challenge*.

In summary, it is clear that all of the analyzed textbooks show improvement for AWL word family repetitions. As such, narrow reading may aid implicit acquisition. Many of the AWL words that increased in frequency were directly related to the topics, but some are more general in nature. The implications of these findings will be discussed further in section 5.2.2.

4.1.4 AWL One Occurrence

There were three parts to the analyses conducted in relation to AWL word families occurring only once. The first analysis examined the number of AWL word families occurring once in each of the 21 texts. The second investigation focused on AWL word families occurring once

across the entire set of topic related texts. Finally, both groups of AWL word families were analyzed with the VP-Complete program to determine their frequency levels.

Table 14. *Percentage of in-text AWL word families used once.*

Textbook and texts	Total AWL word families	1 token	% once
Textbook: Access to English			
Tailored text: Divided by a Common Language	18	14	77,8
Tailored text: A Global Language	34	27	79,4
Authentic text: Renaming English	29	21	72,4
Tailored text: Native Americans – Original Inh.	36	27	75,0
Tailored text: Aboriginal Australians	51	40	78,4
Tailored text: Stolen Children	8	6	75,0
Authentic text: Native Americans In Business	26	17	65,4
Textbook: Stunt			
Tailored text: British vs. American English	17	13	76,5
Tailored text: English as a World Language	18	16	88,9
Authentic text: There Is an Epidemic	47	29	61,7
Tailored text: Native Americans	23	19	82,6
Tailored text: Australia: The Birth of a Nation	17	13	76,5
Tailored text: Stolen Generation	17	11	64,7
Authentic text: Effects of Removal	157	85	54,1
Textbook: Targets			
Tailored text: The Flavours of English	51	39	76,5
Tailored text: The Power of English Part 1	44	33	75,0
Tailored text: The Power of English Part 2	51	34	66,7
Authentic text: English and the Future	48	38	79,2
Tailored text: Native Americans: Still Here	58	44	75,9
Tailored text: Australia – the Island Continent	64	46	71,9
Authentic text: “Indian” mascots	67	51	76,1
Mean score:	881	623	70,7

The overview provided in the above table shows the number of AWL word family in-text occurrences through one token. An average of the total corpus use shows that 70,7% of the AWL word families appeared only one time per text. The rates of frequency range from 54.1% to 88.9%. There is a clear majority of the AWL word families occurring only once per

text in the corpus for this study. The implications of these findings will be discussed further in section 5.2.2.2.

Table 15. *Percentage of AWL word families used once across three and four texts.*

Textbook and Topics	AWL total word families	number of word families once	once %
<i>Access to English</i>			
Global English	70	49	70.0
Indigenous peoples	95	61	64.2
<i>Stunt</i>			
Global English	68	47	69.1
Indigenous peoples	170	87	51.2
<i>Targets</i>			
Global English	139	83	59.7
Indigenous peoples	148	99	66.9
Mean score	690	426	61.7

The number of AWL vocabulary occurring only once in all three or four texts is lowered from an average of 70.7% per text (see table 14) to an average of 61.7 %. However, on average, a majority of the AWL word families used in this corpus still occur as a single token. The rates were also similar between the textbooks. Access to English showed an overall rate of one token per word family across topic related texts at 66.7%, Stunt had the same rate at 56.3% and Targets showed a rate at 63.4%.

Table 16. *AWL word families: BNC/ COCA frequency levels in each textbook, in percentage*

Textbook	< 3,000 high-	3,000 frequency	4,000 - 9,000 mid-frequency
<i>Access to English</i>	42.5	50	7.5
<i>Stunt</i>	38.5	53.3	8.2
<i>Targets</i>	38.9	55.7	5.4

Table 17. *AWL word families: BNC and COCA frequency levels across topics, in percentage*

Textbook	< 3,000 high-	3,000 frequency	4,000 - 9,000 mid-frequency
Access to English	36.4	54.5	10
Stunt	38.1	53.0	9.0
Targets	33.5	61.0	5.5

A second analysis of the BNC/OCA frequency levels for this group of AWL word families was conducted in order to provide more in-depth information concerning what type of AWL word families occurred only once. Calculating the average of each level for each textbook can help explain this trend. As shown in the above tables, the majority of AWL word families (from a low of 50% to a high of 61%) occurring once both in and across texts were at the BNC/COCA 3,000 levels. The possible implications are that, unlike the recycled AWL word families, many of these words may be difficult for L2 learners to understand. Examples of 3,000 level AWL word families used only once in this corpus are: *appreciated, chart, conceived, consequence, diversity, emergence, expansion, functions, identical, illustrate, interpret, neutral, notion, obvious, predict, revise, similar, traced, whereas, brevity, clarity, contributes, diverse, enables, entities, facilitation, furthermore, instance, justifies, label, mutually, nevertheless, precisely, presumably, previously, prospects, and variants.*

4.2 Glossing

As mentioned earlier, glossary items are a form of awareness that can help L2 learners comprehend texts that have a lexical coverage slightly beyond their own vocabulary size (see section 2.4.3.2). The research question guiding this section of the analysis examines to what extent the use of glossaries in tailored texts assist advanced L2 English learners with the acquisition of general academic vocabulary during unassisted reading. Because the research question relates to how glossaries may help L2 vocabulary acquisition, I have let recommendations from researchers also guide the analyses. The glossary analyses have therefore spanned the topics of glossary structure, glossary coverage, AWL use in glossaries and BNC/COCA frequency levels of the AWL glossary items.

Researchers have found that most students prefer glossaries in the margins of the text (Nation, 2013). Since this is the structure used by the materials designers for the analyzed texts, I will not discuss this variable further. All of the glossaries used in this corpus were L2 to L1 translations (see section 3.5.7). Researchers agree that use of L1 translations can be helpful for vocabulary acquisition (see section 2.4.3.2). Only the tailored texts used glossaries so that the authentic texts are not a part of this analysis.

4.2.1 Total Glossary Coverage

The recommended glossary coverage rates are between 3% -5%. Glossary coverage refers to the number of glossary tokens used in each text compared to the total number of tokens in each text (see section 3.5.7). As seen in the table below, glossary coverage, on average, is below the recommended minimum of 3%. The textbook *Targets* is the only book to reach the lowest recommended coverage rate, at 3.1%. Large variations were found between the 15 analyzed texts, from 1.2% to 3.9% glossary coverage (see tables 18-21). This means that there is room for greater use of glossary items in all of the analyzed text related to the current study.

Table 18. Average glossary coverage for the corpus, given in percentage.

Textbooks	Total tokens	Tokens in glossary	% glossary coverage
<i>Access to English</i>	6509	177	2.7
<i>Stunt</i>	3951	87	2.2
<i>Targets</i>	8171	283	3.1
Mean for corpus:	18631	517	2.7

The remaining analyses with regard to glossary use have been conducted in relation to each textbook, in order to provide better detail in relation to the investigated variables.

4.2.1.1 Access to English

In the texts that were analyzed for the textbook *Access to English*, the glossaries made up 2.7% of the total text, on average. This is slightly below the recommended glossary coverage of 3% - 5% (Nation, 2013).

Table 19. Glossary coverage for tailored texts in *Access to English*, in percentage.

Tailored texts	Total tokens	Tokens in glossary	% glossary coverage
Divided by a Common Language	1066	13	1.2
A Global Language	2076	33	1.6
Aboriginal Australians	1705	79	4.6
Stolen Children	297	8	2.7
Native Americans - Original Inhabitants	1365	44	3.2
Mean score	6509	177	2.7

4.2.1.2 *Stunt*

On average, 1.3% of the analyzed texts in *Stunt* were used in a glossary, clearly below the recommended 3-5% glossary coverage (Nation, 2013).

Table 20. *Glossary coverage for tailored texts in Stunt, in percentage.*

Tailored texts	Total tokens	Tokens in glossary	% glossary coverage
British vs. American English	1114	32	2.9
English as a World Language	628	10	1.6
Native Americans	900	16	1.8
Australia: The Birth of a Nation	848	16	1.9
Stolen Generation	461	13	2.8
Mean score	3951	87	2.2

4.1.1.3 *Targets*

On average, 2.7 % of the analyzed texts in *Targets* were used in a glossary, slightly below the recommended 3-5% glossary coverage (Nation, 2013).

Table 21. *Glossary coverage for tailored texts in Targets, in percentage.*

Tailored texts	Total tokens	Tokens in glossary	% glossary coverage
The Flavours of English	2187	42	1.9
The Power of English Part 1	1161	35	3.0
The Power of English Part 2	1866	73	3.9
Native Americans: We Are Still Here	1440	47	3.3
Australia – the Island Continent	1517	56	3.7
Mean score	8171	253	3.1

4.2.2 AWL Glossary Coverage

The glossary coverage of AWL word families has been analyzed in several different ways. The first findings show how many of the glossary items used in the 15 texts were on the AWL. The second set of findings describes a closer analysis of the AWL word families that were glossed. For these analyses AWL word families that occurred once in the text were investigated in relation to glossing and compared to the total number of AWL word families used once in the tailored texts. A Range analysis was not conducted because only tailored texts contained glossaries.

4.2.2.1 Total AWL Glossary Coverage

The following table describes how many of the glossary items were AWL word families. This is compared to the total number of AWL word families used in each text. In this manner, it is possible to show how much AWL vocabulary is glossed, both in relation to other glossary items and the number of AWL word families used in each text as a whole.

Table 22. Total AWL glossary coverage per textbook , in percentages.

Textbook and tailored texts	AWL glossary coverage	In-text AWL word families
<i>Access to English</i>		
Divided by a Common Language	27.3	16.7
A Global Language	20.5	23.5
Aboriginal Australians	19.4	23.5
Stolen Children	14.3	12.5
Native Americans: Original Inh.	24.3	25
Mean score:	21.2	20.2
<i>Stunt</i>		
British vs. American English	15.8	27.3
Global English	23.3	55.6
Native Americans	12.5	30.0
Tailored text: Australia: The Birth of a Nation	0.0	0.0
Tailored text: Stolen Generation	15.8	27.3
Mean score:	13.5	28.0
<i>Targets</i>		
The Flavours of English	10.2	24.0
The Power of English Part 1	24.0	35.3
The Power of English Part 2	10.6	11.9
Native Americans: Still Here	13.8	25.0
Australia – The Island Continent	12.0	17.5
Mean score:	14.1	22.7

On average, 21.2% of the glossary terms in *Access to English* were AWL word types. The glossed AWL types represent, on average 20.2% of the entire amount of AWL vocabulary used in the texts. A VP-Compleat analysis of the glossary terms show that over half of the glossary terms were within the high-frequency vocabulary range (see Appendix 7.3.1). Seen as a whole, there were more glossary terms from the first 2,000 word frequency levels than the 3,000 frequency levels. It may be assumed that L2 learners may have a good understanding of the more frequent words so this finding was surprising. There is a larger average percent coverage of mid-frequency vocabulary, as would be expected. Any remaining vocabulary items were either from the off-list category or from the low frequency range above 9,000.

An average of 13.5% AWL word types were represented among the total number of glossary terms for each text represented in *Stunt*. However, large variations between texts in relation to this variable were found. One text included 55.6% of the total in-text AWL vocabulary in the glossary and one text had no AWL types in the glossary. On average, 28% of the AWL vocabulary used within each text has been translated in the glossary. An average of 32% of the glossary items are in the first 2,000 range, while 31% are at the 3,000 level. An average of 35% are mid-frequency vocabulary. The remaining glossary items are either above the 9000 frequency level or are on the off-list category (see Appendix 7.3.2). There is a slightly higher use of glossary translations for mid-frequency vocabulary types, which is the recommended use for glossaries. However, over one third of the terms defined in the glossaries in this text are high-frequency items.

Looking at all of the texts analyzed in *Target* as a whole, nearly 23 % of the vocabulary used in the analyzed texts from this textbook were AWL types. An average of slightly over 14% of the AWL word types used in the text were translated in the glossary. Seen as a whole, an average of 26.4% of the glossary items are in the high-frequency range, while 34.2% are mid-frequency vocabulary. An average of 16.9% of the glossary items are used to translate collocations. The remaining glossary items are either above the 9000 frequency level or are on the off-list category (see Appendix 7.3.3). There is a slightly higher use of glossary translations for mid-frequency vocabulary types, which is the recommended use for glossaries. However, over one quarter of the terms defined in the glossaries in this text are words deemed to be high-frequency words. Of these an average of 24.6% were found in the first 2,000 frequency levels while an average of 28.3% were at the 3,000 frequency level.

As expected, a minority of AWL word families were represented in the text glossaries. The rates of coverage varied from zero to 27%. Differences between textbooks show that *Access to English* had the largest coverage rate at 21.2%. *Stunt* and *Targets* had quite similar rates at 13.5% and 14.1% respectively. Many of the glossary items in all of the textbooks were high-frequency word families.

At the same time, it is clear that only a small minority of the AWL word families used in the entire text are defined in the glossaries. There were small differences between textbooks. A larger portion of the AWL word families found in *Stunt* were defined in the glossary, 28%. *Access to English* and *Targets* had similar values at 20.2% and 22.7%. In other words, on average over 70% of the AWL word families used in the texts were not found in the glossaries.

4.2.2.2 Glossed AWL with one in-text occurrence

The following analyses were conducted in order to examine if AWL word families only occurring once in a text would be included in glossaries as this might help L2 acquisition.

Table 23. *Per textbook: AWL glossary coverage with one occurrence, in percentages.*

Textbook and tailored texts	Glossed AWL used once	AWL used once in text
<i>Access to English</i>		
Divided by a Common Lang.	66.7	14.3
A Global Language	87.5	25.9
Aboriginal Australians	66.7	20
Stolen Children	100	16.7
Native Americans: Original Inh.	66.7	18.8
Mean score:	77.5	19.1
<i>Stunt</i>		
British vs. American English	33.3	7.7
Global English	60	18.9
Native Americans	33.3	5.3
Australia: The Birth of a Nation	0.0	0.0
Stolen Generation	100.0	27.3

Textbook and tailored texts	Glossed AWL used once	AWL used once in text
Mean score:	45.3	11.8
<i>Targets</i>		
Tailored text: The Flavours of English	66.7	10.3
Tailored text: The Power of English Part 1	83.3	30.3
Tailored text: The Power of English Part 2	50	8.8
Tailored text: Native Americans: Still Here	70	15.9
Tailored text: Australia – The Island Continent	57.1	8.7
Mean score:	65.4	14.8

There was large variation between textbooks when it comes to how many of the glossed AWL word families were only used once in the text. On average, a slight minority (45.3%) of the AWL glossed items occurred once in the text representing *Stunt*. In *Targets*, an average of 65.4% of these items occurred once in the texts. *Access to English* had the highest rate (77.5%) of AWL word families only used once in the text that appeared as glossary items. This means that in two textbooks a slight majority of the AWL glossary terms can be an extra help for students' acquisition of general academic vocabulary, since none of the terms had a frequency of six or more occurrences.

However, a relatively small percentage of the AWL types used once in the 15 texts were found among the glossary terms. In *Stunt*, just under 12% of the AWL word families used in the text were represented in the glossary. The rates for *Target* and *Access to English* were 14.8% and 19.1% respectively. This means that at a minimum, 80% of the AWL word families used in the tailored texts studied here were used once in the text were also not defined in the glossary.

A closer description of each textbook provides better detail of what this finding means. In *Access to English*, over 75% of the AWL vocabulary used once on the text were not included in the glossary. Some examples of these AWL terms not glossed are: *context, norm, contrary, variants, contributed, displacement, fundamentally, traces, apparent, decline, estimated, phenomenon, and precision*. In nine out of the above thirteen AWL tokens, just over 80%, were on the 3,000 and 4,000 frequency levels. The text book *Stunt* had wide

variation between texts in relation to the number of word families represented in the glossary that were only used once in the text, from 100% to 33.3%. On average, 11.8% of the total number of AWL word families used once in the texts were translated in the glossaries. A high percentage of AWL types, on average 88.2%, occurring only once in the texts are not included in the glossary. Some examples of terms used once that are not glossed are *distinct, instance, established, ensured, dominant, require, alter, estimate, primary, perspective, resource, behalf, construct, seek, and site*. Of the examples provided here, 60% are above the 2,000 frequency level. In *Targets*, a high percentage of AWL types, on average 85.2%, occur only once in the texts without glossary support for L2 readers. Some examples of AWL terms used once without being glossed in this textbook are: *chart, conceived, conflicting, interpret, revise, academics, brevity, clarity, facilitation, variants, contribute, ideological, and visible*. Of the above examples, 84.6% are found at the 3,000 frequency level or over. Data for the frequency levels were found through VP-Compleat analyses.

4.3 Lexical Coverage

The following section will first present a brief discussion related to findings concerning the total vocabulary use in the corpus of 21 factual texts. The analyses have investigated reading comprehension measured in terms of lexical coverage and vocabulary size (see section 2.4.4). These analyses have been conducted in order to discuss more completely the AWL vocabulary and the prospects of implicit AWL vocabulary acquisition for L2 learners during unassisted reading in light of the total vocabulary use found in these texts. The following analyses show what vocabulary size L2 learners need to reach 98% and 95% lexical coverage for texts in this corpus.

4.3.1 General Lexical Coverage

Most researchers today agree that to comprehend a text appropriately, L2 learners should have knowledge of around 98% of the vocabulary used in a text; 95% lexical coverage is seen as a minimum (Laufer, 2010; Nation, 2013; N. Schmitt et al., 2015). Studies show that to comprehend authentic texts for general English L2 learners may need a vocabulary size as high as the 8,000 - 9,000 frequency levels (Nation, 2006; D. Schmitt & Schmitt, 2012).

In the following, the general vocabulary size needed to reach a lexical coverage of 98% will be presented first. This will be followed by a brief presentation of vocabulary size needed for 95% lexical coverage. Data for the statistics presented here can be found in the VP-Compleat analyses in Appendix 7.2.

Table 24. *BNC and COCA frequency levels entire corpus, in percentage of total tokens*

Textbook and texts	Propers + < 3,000	3,000	4,000 – 9,000
Textbook: Access to English	High- frequency	Mid-frequency	
Tailored text: Divided by a Common Language	90.76	3.45 (94)	3.17 (97)
Tailored text: A Global Language	92.24	3.90 (96)	2.42 (99)
Authentic text: Renaming English	88.86	7.53 (96)	2.67 (99)
Tailored text: Native Americans: Original Inhabitants	90.47	5.79 (96)	2.19 (99)
Tailored text: Aboriginal Australians	89.97	4.69 (94)	3.29 (98)
Tailored text: Stolen Children	92.93	4.38 (97)	2.35 (100)
Authentic text: Native Americans In Business	90.35	5.63 (96)	1.76 (100)
Textbook: Stunt			
Tailored text: British vs. American English	91.23	2.51 (94)	4.48 (98)
Tailored text: English as a World Language	94.17	3.18 (97)	1.23 (99)
Authentic text: There Is an Epidemic	89.89	5.97 (96)	3.09 (99)
Tailored text: Native Americans	90.78	4.78 (96)	2.77 (98)
Tailored text: Australia: The Birth of a Nation	92.80	4.72 (96)	2.24 (99)
Tailored text: Stolen Generation	93.06	4.56 (98)	2.18 (99)
Authentic text: Effects of Removal	87.75	7.55 (95)	3.47 (99)
Textbook: Targets			
Tailored text: The Flavours of English	90.57	3.71 (94)	2.65 (97)
Tailored text: The Power of English Part 1	90.26	7.15 (97)	1.12 (99)

Tailored text: The Power of English Part 2	90.51	6.11 (97)	2.20 (99)
Authentic text: English and the Future	92.52	5.58 (98)	1.44 (100)
Tailored text: Native Americans: We Are Still Here	86.95	6.88 (94)	3.17 (98)
Tailored text: Australia – the Island Continent	85.33	5.91 (91)	3.98 (95)
Authentic text: “Indian” Mascots	85.21	7.46 (93)	4.22 (97)

As shown in the diagram below and table 24, in order to achieve 98% lexical coverage rates, L2 learners would need a vocabulary size up to the first 9,000 frequency levels in a majority of the texts analyzed in this corpus (66.7%)⁶. Nearly 20% of the texts (19%) did not provide 98% lexical coverage even with this vocabulary size. Three texts, one authentic and two tailored, provided lexical coverage rates of 98% with a vocabulary size of high-frequency vocabulary, however. Without help to understand mid-frequency vocabulary, many texts may be difficult for advanced L2 learners to comprehend (see section 5.4.1).

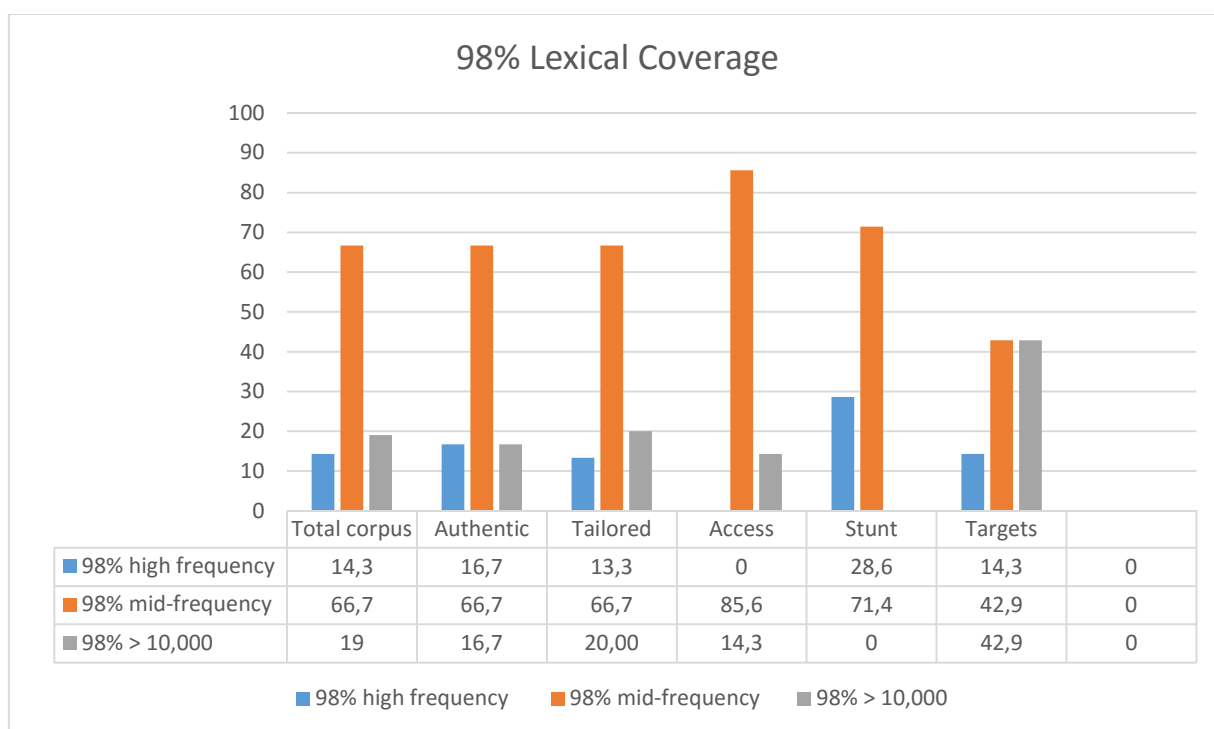


Figure 3. Lexical coverage in relation to vocabulary size, in percentage

⁶ High-frequency vocabulary is defined using categorization of BNC/COCA frequency levels of 1,000 – 3,000 and mid-frequency vocabulary represents word families from the 4,000 to 9,000 levels. These are the definitions presented by Schmitt and Schmitt (2012) replacing the prior vocabulary goals of GSL and AWL vocabulary (see section 2.4.1).

However, as shown in the following table, a majority of the texts (71.4%) did provide 95% lexical coverage with vocabulary knowledge of high-frequency word families. In relation to the current study, it is important to keep in mind these texts are a part of course materials and that they, in many cases, will be read under supervision of a teacher. In this classroom situation, L2 learners do not only engage in unassisted reading and therefore it has been important to analyze these rates of lexical coverage as well.

4.3.1.1 *Authentic vs. Tailored texts*

Clear differences between authentic and tailored texts were found in the corpus used for the current study. A slightly greater number of tailored texts failed to provide 98% lexical coverage at mid-frequency levels of vocabulary, 20% versus 16.7% respectively. At the same time, a larger number of the authentic texts reached 95% lexical coverage with high-frequency level vocabulary size, 83.3% of the authentic texts versus 66.7% of the tailored texts. This may mean that the authentic texts can be easier to comprehend, even though they used more academic vocabulary.

Upon closer study, it is possible to see an interesting trend in relation to general vocabulary use in the corpus gathered for the current study. There is a larger ratio between the use of mid-frequency vocabulary and AWL vocabulary in the authentic texts, compared to tailored texts (see Appendix 7.7). All of the authentic texts use relatively more AWL vocabulary than mid-frequency vocabulary. This is important because it can mean that these texts are easier for L2 learners to comprehend even though they use larger amounts of AWL vocabulary. In three of the tailored texts, there was a greater use of mid-frequency vocabulary than AWL vocabulary.

A more in-depth study of one such text explains the situation better. In the tailored text “A Global Language” there are 39 word families in the mid-frequency range, not including three AWL families in this range. For the same text 18 AWL word families are used, three of which are at 4,000 and 5,000 frequency levels. Seven of the 39 mid-frequency words are translated in the glossary (see Appendix 7.2.1.2). The glossary coverage rate for this text was 1.6% (see table.). One of the AWL word families, *globe*, was recycled enough in the Range analysis to promote implicit acquisition. The other two AWL word families *contrary* and *immigrate* were used once. Examples of mid-frequency vocabulary that was not glossed and

occurred once in the text includes the word types *departure*, *destination*, *descent*, *foremost*, and *tentative*. Thirty-two mid-frequency words were not defined in the glossary and of these, 75% were used only once in the text.

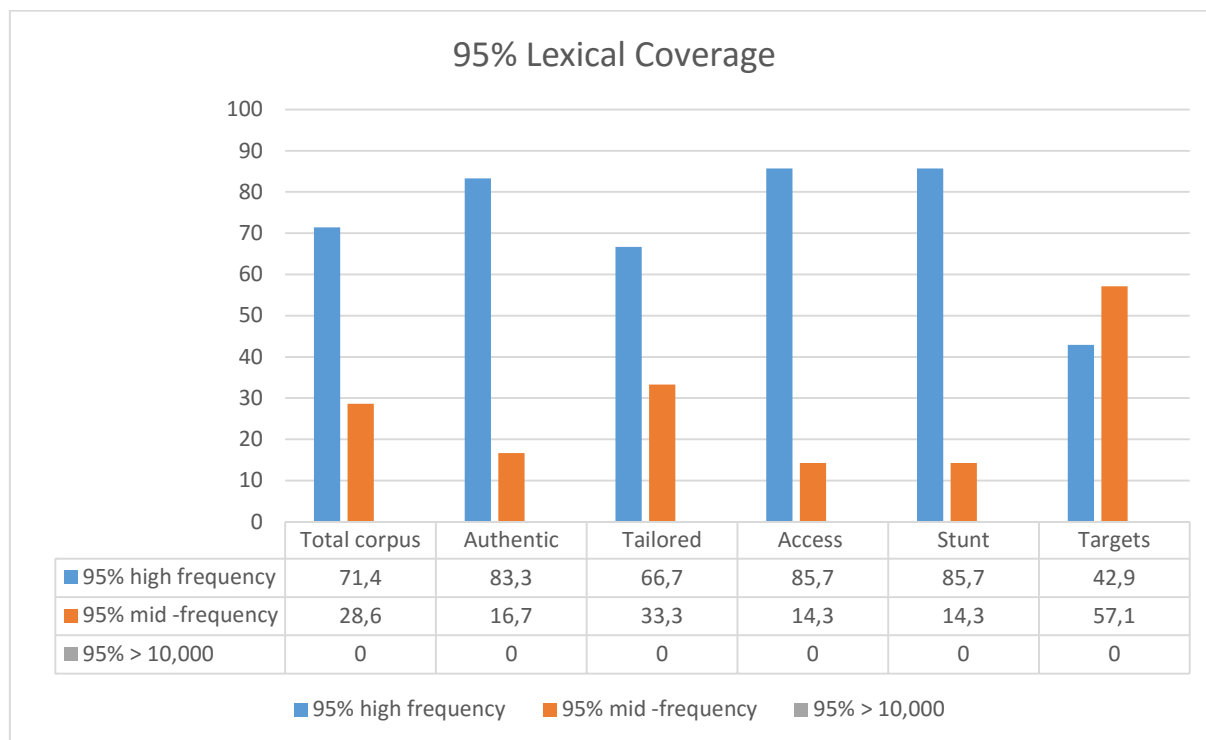


Figure 4. *Lexical coverage in relation to vocabulary size, in percentage*

4.3.1.1 Textbook comparison

As shown in the above diagrams, there is large variation in lexical coverage between the different textbooks. For the textbook *Access to English*, to gain 98% lexical coverage, L2 learners must have form-meaning, receptive knowledge of both high and mid-frequency vocabulary for five of the seven texts (66.7%). Three texts, one authentic and two tailored, provided a lexical coverage of 98% with a high-frequency vocabulary, however. lexical coverage of 98% until learners have vocabulary knowledge above the 9000 frequency level, but one of them does provide 95% coverage at the high-frequency level.

In *Stunt*, all of the texts provide at least 98 % lexical coverage for L2 learners with high and mid-frequency vocabulary knowledge. Six out of seven texts (85.7%) analyzed in this textbook provided 95% lexical coverage for students who have a vocabulary size covering high-frequency vocabulary.

For the textbook *Targets*, four out of seven texts (57%) require vocabulary knowledge above the 9,000 frequency level for 98% lexical coverage. Three out of seven texts (42.9%) analyzed in *Targets* provide 95% lexical coverage for students who have a vocabulary size covering high and mid-frequency vocabulary.

Both *Access to English* and *Stunt* provide a clear majority of texts that L2 learners with a vocabulary size up to the 9,000 frequency level may comprehend adequately during unassisted reading. The lexical coverage in *Targets* places much higher demands on L2 learner vocabulary size. If glossary terms are not used to aid comprehension of mid-frequency vocabulary, students with an understanding of high and mid-frequency vocabulary may not understand a majority of the texts adequately.

4.3.2 AWL Lexical Coverage

As mentioned earlier, the word families on the AWL largely represent high-frequency vocabulary (see section 2.1.1). This means that, perhaps contrary to popular belief, texts using general academic vocabulary are not necessarily difficult for L2 learners to comprehend. One other aspect of the current study has been to analyze the AWL vocabulary used in the entire corpus also in relation to BNC and COCA frequency levels. This has been done in order to examine the expected learning burden these vocabulary words may pose. Since there is large variation between texts and textbooks for these variables the analysis will be presented per textbook.

4.3.2.1 Access to English

As could be expected, a clear majority of the AWL vocabulary used in the analyzed texts had high frequency levels. In five out of seven texts, 50% or more of the AWL vocabulary used are found at the 1,000 and 2,000 frequency levels. In all of the texts, the AWL vocabulary used is found at high-frequency levels. A majority of the AWL vocabulary used in two of the texts has a frequency level of 3,000. All of the seven texts use AWL vocabulary in the mid-frequency range. On average, the mid-frequency coverage for this vocabulary use is at 8%. This is the highest level of mid-frequency AWL vocabulary use of the three textbooks. At the same time, *Access to English* also had the lowest percentage of AWL vocabulary use in this study at an average of 4,8% (see tables 3 and 4).

Table 25. *Access to English: BNC/COCA frequency levels of AWL word families, in percent*

Textbook and texts	K1- K2 Propers + < 3,000	K3 3,000	K4- K9 4,000 –9,000
<i>Access to English</i>	High-frequency		Mid-freq
Tailored text: Divided by a Common Language	55.56	38.89	5.56
Tailored text: A Global Language	32.35	52.94	14.71
Authentic text: Renaming English	27.59	65.52	6.9
Tailored text: Native Americans: Original Inh.	61.11	36.11	2.77
Tailored text: Aboriginal Australians	54.90	43.14	2
Tailored text: Stolen Children	50.00	37.5	12.5
Authentic text: Native Americans In Business	50.00	38.46	11.54

4.3.2.2 *Stunt*

An equal number of texts, three of seven (42.9%), used a majority of AWL vocabulary at the 1,000 and 2,000 frequency level and at the 3,000 level. Four of seven texts (57.1%) made use of AWL vocabulary in the mid-frequency range. For these texts the average of mid-frequency level vocabulary was 7.2%. There is, on average, a majority of AWL vocabulary from the first 2,000 frequency levels used in *Stunt*, but this seems largely due to a wide diversity between texts, more than the fact that one group of the high-frequency ranges is used constantly more than another. The same applies to the use of AWL vocabulary in the mid-frequency range. This makes an overall judgment of AWL use in relation to vocabulary size difficult. What can be said, is that the two authentic texts are quite similar, with a majority of AWL vocabulary at the 3,000 level and a little more than 6% in the mid-frequency range. The tailored texts vary widely from 65% in the first 2,000 range to 58% in the 3,000 range. It would seem that L2 learner comprehension may vary greatly from text to text. Three tailored texts do not use AWL in the mid-frequency range, but one of them, “Australia – Birth of a Nation”, has over 11% use of mid-frequency AWL vocabulary, in a text with only 2.6% AWL vocabulary coverage in total. The two mid-frequency word types in this text are *behalf* and *immigrate*. They are used once and twice in the text respectively and are not included in the glossary.

Table 26. *Stunt: BNC/COCA frequency levels of AWL word families, in percent*

Textbook and texts	K1- K2 Proprs + < 3,000	K3 3,000	K4- K9 4,000 –9,000
<i>Stunt</i>	High-frequency		Mid-freq
Tailored text: British vs. American English	41.18	58.82	0
Tailored text: English as a World Language	55.56	44.44	0
Authentic text: There Is an Epidemic	42.55	51.06	6.38
Tailored text: Native Americans	65.22	30.43	4.35
Tailored text: Australia: The Birth of a Nation	41.18	47.06	11.76
Tailored text: Stolen Generation	64.71	35.29	0
Authentic text: Effects of Removal	40.76	52.87	6.37

4.3.2.3 Targets

Fifty percent or more of the AWL vocabulary used in two of the seven texts (28.6%) were at the 1,000 and 2,000 frequency levels. Three of the seven texts (42.9%) used a majority of the AWL vocabulary at the 3,000 frequency level. All of the seven texts use AWL vocabulary in the mid-frequency range, on average, 4.8%. Though *Targets* has the highest average percentage use of AWL vocabulary of the three analyzed textbooks, the use of mid-frequency level AWL is lowest. The division between the two high-frequency AWL vocabulary groups was nearly equal.

Table 27. *Targets: BNC/COCA frequency levels of AWL word families, in percent*

Textbook and texts	K1- K2 Proprs + < 3,000	K3 3,000	K4- K9 4,000 –9,000
<i>Targets</i>	High-frequency		Mid-freq
Tailored text: The Flavours of English	50.98	45.1	3.92
Tailored text: The Power of English Part 1	38.64	56.82	4.55
Tailored text: The Power of English Part 2	56.86	37.25	5.88
Authentic text: English and the Future	45.83	52.08	2.08
Tailored text: Native Americans: We Are Still Here	44.83	50.00	5.17
Tailored text: Australia – the Island Continent	45.31	48.44	6.25
Authentic text: “Indian” Mascots	47.76	46.27	5.97

A closer examination of the authentic text “English and the Future” shows that though 7% of the vocabulary was found on the AWL, the only mid-frequency AWL word family used for this text was *globe*. The term was used four times in the text, but the text did not include a glossary. It should be pointed out that a majority of the AWL vocabulary (52%) was at the 3,000 frequency level and, since there is no glossary, some advanced L2 learners may still have comprehension difficulties. In comparison, the authentic texts studied more closely did not include the mid-frequency AWL word families in the glossaries and these words were only used once or twice in the text.

4.3.3 In-depth Investigation of One Text

I have chosen the text “The Power of English Part 2”, to provide a more in-depth investigation of glossary use because the highest glossary coverage level at 3.9% was used for this text. The investigation analyzed glossary use in relation to lexical coverage and vocabulary size represented in the text, as well as in relation to AWL vocabulary glossed.

In order to reach the recommended 98% lexical coverage rate for this text, students must have form-meaning, receptive knowledge of high-frequency and mid-frequency level vocabulary. If students do not know this vocabulary, the translation of glossary terms may help better their text comprehension (Nation, 2013). An analysis of mid-frequency level vocabulary shows that 56 % have been translated in the glossary, and all of the low-frequency types were included in the glossary. Mid-frequency vocabulary not translated included such words as *emigration*, *consolidation*, and *ideological*.

High-frequency vocabulary contains the highest rates of vocabulary frequency, and thus advanced L2 students should have greater word knowledge of these word families. Nearly half (49.1%) of the glossary terms are found at this frequency level. Three words, *far-reaching*, *post*, and *settlement* are found at the 1000 frequency level and may be terms many advanced L2 learners know. However, the word *post*, has different meanings and in this context it is understandable that this word was glossed. Another nine terms in the glossary are found at the 2,000 frequency level and are also word types these learners should be familiar with. Examples from this group include *claim*, *exposed*, *gain*, *trade*, and *root*. Vocabulary items at the 3,000 word level are also considered high-frequency, but may represent some difficulties for advanced L2 learners. Of the terms from this level, 21.9% are represented in

the glossary. Three thousand frequency level words that are not glossed include *administrative, formalized, founded, imperial, and launched*.

With regard to the use of AWL vocabulary in “The Power of English Part 2”, there is a relatively high percentage (5.2%) of overall AWL vocabulary use. Of the 51 AWL word families represented in the text, 34 (66.7%) occur once. Three of these word families occur six or more times in the text, *culture, economy* and *immigrate*. None of these word types are defined in the glossary, which is positive as the recycling of these terms in themselves should help acquisition if they are unknown to the reader. Of the AWL glossary items for this text three of the six (50%) AWL types were used only once in the text. Two AWL types were used two and three times in the text, clearly under the recommended minimum of six repetitions for implicit acquisition. The final AWL type *enforce*, was not used in the text, but is included in the glossary. Four of the six AWL types (66.7%) included in the glossary are also at the 3,000 frequency level. The use of these terms in the glossary would be expected to help learners acquire these AWL word families.

At the same time, only 8.8% of the AWL vocabulary used once in this text were included in the glossary. Of the remaining AWL word families not included in the glossary, twenty were at the 3,000 frequency level or higher. Examples of AWL tokens at the 3,000 level only used once in the text and not included in the glossary are *aspects, founded, ideological, and transformed* (see section 7.2.3).

A majority of the glossary terms in the text would aid students in raising their comprehension, thus making the text more comprehensible for advanced L2 learners with a smaller vocabulary size. At the same time, there are also a number of words included in the glossary that advanced L2 learners can be expected to know. The glossary for this text had a clear minority of AWL vocabulary. Many of the AWL word families also only occur once in the text. That being said, textbook authors must focus on the content of a texts as well as the vocabulary used to relay the desired information.

5. Discussion of Results

The following chapter will provide a discussion of findings related to the analyses presented in chapter four. These will be compared to prior research and placed within a theoretical framework. Section 5.1 briefly recaps the aims of the current study, as well as the research methodology applied. A discussion of findings relevant to AWL vocabulary usage in the analyzed texts will be presented in section 5.2. Section 5.3 provides an examination of the findings related to AWL vocabulary used in glossaries. Finally, in section 5.4 findings related to lexical coverage and vocabulary size will be discussed.

5.1 Brief Overview

The aim of the current study has been to investigate the use of general academic vocabulary in textbooks used in obligatory, college preparatory English courses for Norwegian high school students. In doing so, the study aims to provide a better understanding of how general academic vocabulary is used in course materials and help bring research connected to general academic vocabulary into a Norwegian context. It is hoped that the study will lead to a better understanding of the extent to which this vocabulary use provides the means for the implicit acquisition of these words.

Forming the basis of the current study is the understanding of AWL items as well-documented examples of general academic vocabulary (see section 2.1). Usage-based theory related to vocabulary acquisition and relevant hypotheses have been used as a theoretical framework. The research questions guiding the analyses conducted here are focused on AWL vocabulary usage, glossing and lexical coverage.

Research for the current study has been conducted using mixed methods. A study corpus was formed from 21 texts in three different English subject textbooks. A total of 28,734 tokens made up the vocabulary in this corpus. Quantitative methods related to corpus linguistics were used when gathering numeric data. In-text and across text vocabulary use has been sorted according to frequency and general frequency levels with the aid of computer programs. The use of qualitative research methods provided further information about the numeric data found in the quantitative research. The corpus created for the current study is small in size, allowing for an in-depth discussion of general academic vocabulary use in texts and textbooks. Parts of the data collection process have been conducted manually, when

computer programs did not provide needed measures. The gathered information has also been manually sorted and organized into diagrams and tables.

5.2 AWL Vocabulary Use

AWL coverage analyses and frequency studies have been applied to the corpus in order to help answer research question 1a pertaining to AWL word family usage in the corpus (see section 1.4). The findings from analyses related to AWL coverage rates and frequency will be discussed in the following.

5.2.1 AWL Text Coverage

AWL coverage refers to the overall percentage of vocabulary in each text that contained AWL word families (see section 3.5.3). These coverage rates were seen in relation to authentic genre comparable to the factual texts represented in course materials i.e., authentic English language newspapers (ca 5%) and academic texts used in university studies (8%-10%) (Coxhead, 2000; Nation, 2013). The findings presented in section 4.1.1 showed that a slight majority of the texts provided an AWL coverage rate of 5% or more. Even so, over 45% of the texts showed coverage rates below 5%. Only two texts provided 8% coverage rates. In other words, despite being factual texts, the majority of texts in this corpus (85%) would not be comparable to academic texts and many of these factual texts used less AWL vocabulary than would be expected in English language newspapers. If L2 students are to learn academic vocabulary implicitly, they will need large amounts of exposure to these words (Cobb, 2007; Krashen, 2013; Nation, 2013). In usage-based theory, L2 learners are dependent upon usage events (in this case, encounters with words) in order to develop even the most basic form of word knowledge i.e., form-meaning, receptive knowledge (see section 2.4.2.2). It is only through a repetition of encounters with symbolic units, such as words, that these form-meaning connections can develop into vocabulary acquisition (N. Schmitt & Verspoor, 2013).

In relation to AWL coverage, there were large differences found between the authentic and tailored texts. Authentic texts had larger rates of coverage than tailored texts, on average, 7.5% versus 4.1% (see table 1). These findings suggest that if only tailored texts are used in the classroom, exposure to AWL vocabulary may be reduced considerably. The diversity of contexts in which usage events take place is also important for the process of association and

schematization. In order to develop better word knowledge usage-based theory contends that learners will need to experience words being used in different ways. This lends strong support for the use of authentic texts in classroom settings (Langacker, 2000; N. Schmitt & Verspoor, 2013). Through the use of authentic texts, in combination with textbook texts, L2 learners may not only gain a larger percentage of AWL exposure, but will also experience AWL vocabulary in a wider range of contexts. This may also help students gain a more abstract understanding of the AWL vocabulary thus increasing their word knowledge.

That being said, the textbooks are written for first year high school students, so that AWL rates of coverage should perhaps not be compared to university materials. Coverage rates around 5% may be a good starting point, as newspapers could be a very important source for further academic vocabulary development (Nation, 2013). However, the fact remains that these English courses are the last obligatory courses at the high school level in Norway. As such, they are also the last English course many students will take before entering college studies. With low rates of exposure to general academic vocabulary in these textbooks, students may experience a very wide gap between high school and college texts. Also, when recent studies show that the AWL word families are among some of the most frequently used vocabulary items in authentic English discourse (Cobb, 2010) it is perhaps necessary to reevaluate the use of these terms in course materials.

Another important finding from the current study in relation to AWL coverage is that, despite the relatively low use of AWL coverage, the texts containing a small percentage of AWL vocabulary are not necessarily easier for advanced L2 learners to comprehend. A more in-depth discussion of this finding is provided in section 5.4.2.

There are many research studies concerning the use of AWL vocabulary in university texts, but few that examine course materials for advanced L2 learners at the high school level. The findings from the current study partly support findings from a recent study conducted for the Akita International University in Japan. Researchers found that the vocabulary used in these textbook texts did not necessarily reflect the target group for these books. This despite the fact that they were textbooks widely used on the international market (Ruegg & Brown, 2014). This quantitative study by Ruegg and Brown (2014) analyzed vocabulary use in one text from 20 different English as a Second Language (ESL) textbooks for learners at different proficiency levels. Several books written for upper-intermediate L2 learners had an overuse of words from the 1,000 GLS level. As such, they question the pedagogical appropriateness of the vocabulary used in some of the texts for their study (Ruegg & Brown, 2014). While the

current findings do not put into question the vocabulary appropriateness of the analyzed textbook texts, they do suggest that AWL exposure will be limited with only the use of tailored texts.

After conducting several Norwegian studies of high school and college students' English reading proficiency Hellekjær questions the appropriateness of English elective course materials. He claims that the texts are "too often at a language...level that provides little or no challenge for the students (My translation from 2012a, p. 31). It has been the aim of this study to help provide more knowledge in this area, though I have chosen to assess textbooks related to obligatory English courses. My findings suggest that the vocabulary used is challenging enough (see section 5.4), but there are often low rates of exposure to the general academic vocabulary students will need for university level studies.

According to usage-based theory, implicit acquisition of general academic vocabulary is dependent upon large amounts of exposure in many different settings (Langacker, 2000; N. Schmitt & Verspoor, 2013). The findings from this section of the study show that, though the overall use of AWL vocabulary is relatively low, the use of authentic texts may help L2 learners in both areas.

5.2.2 Range Frequency

Another way in which exposure to AWL vocabulary in factual, textbook texts has been analyzed in the current study is by assessing how often the AWL word families presented in the corpus were repeated. This vocabulary recycling was measured for each text and across topic related texts, in order to examine more closely if AWL vocabulary use could be expected to promote implicit vocabulary acquisition.

5.2.2.1 Frequency of six or more repetitions

Findings from the in-text frequency for AWL vocabulary showed that there were few AWL word families repeated enough to promote implicit acquisition. In a corpus of 28,734, only 2.6% of the AWL word families used were repeated the desired six times or more. A majority of the 21 analyzed texts did not contain any words that were recycled enough to promote unassisted learning. The Frequency Hypothesis postulates that what parts of a second language are learned first is dependent upon how often they occur (Hatch & Wagner-Gough (1976) in R. Ellis, 2008). The findings in the current study clearly show that a majority of the academic vocabulary used here would not be among words students would learn first during

unassisted reading. Another interesting finding was that a majority of the recycled AWL words were among the first 2,000 BNC/COCA frequency levels and would likely represent words students know. This is an important finding because L2 learners at the level of proficiency expected of first year high school students in Norway should already be familiar with vocabulary up to the 2,000 frequency level (Nation, 2013). *Targets* was the only textbook that had a majority of texts recycling AWL word families (ca 70%). These findings imply that students using these textbooks could be expected to have prior knowledge of most of the recycled AWL word families present in the analyzed texts. In other words, it would seem that very few AWL word families may be learned implicitly through unassisted reading and as such, explicit vocabulary instruction for AWL word families would be recommended.

Zipf's law (see section 2.4.2.1) explains this tendency in mathematical terms. There are many words in the English language that are not used very often, but a very few words, around the first 2,000 -3,000 frequency levels, are used very often (Nation, 2013). The implications of this law are also shown in findings related to narrow reading and AWL only occurring once.

5.2.2.2 Narrow reading

Reading several topic specific texts together i.e., narrow reading, did enhance vocabulary recycling of AWL vocabulary, though gains were relatively small (see section 4.1.3). Again, a majority of the recycled words were at the first 2,000 BNC/COCA frequency levels and it could be expected that advanced L2 learners have prior knowledge of these words.

5.2.2.3 AWL word families occurring once

In order to conduct a more in-depth investigation into AWL word family usage, the word families used once in the corpus were also examined, in relation to both in-text frequency and the Range analyses. The findings showed that a clear majority of the AWL word families appeared only once (see section 4.1.4). Unlike the recycled AWL word families, a majority of the words used once are found at the BNC/COCA 3,000 frequency level. The possible implications of this are that, unlike the recycled AWL word families, many of these words may be difficult for L2 learners to understand. This being the case, all of the findings related to frequency of AWL vocabulary use suggest that it is unlikely for students to acquire knowledge of a majority of the AWL vocabulary used in this corpus without some form of

instruction. More importantly, it is the vocabulary expected to be outside the vocabulary size of advanced L2 learners that occurs least, again following the laws of vocabulary distribution.

One implication of Zipf's law may be the importance for teachers to recognize that it will take a lot of reading to increase advanced L2 students' vocabulary size implicitly (Cobb, 2007; Krashen, 2013). The current study shows that in order to acquire this vocabulary implicitly, students will need to read much more than the factual texts included in these textbooks. For many students, vocabulary instruction in relation to general academic vocabulary is something that will be needed at all levels of study.

5.2.2.4 Previous studies

The findings from this study are supported by previous research showing that, from the 3,000 frequency level, much fewer repetitions will occur within a text (Cobb, 2007; Matsuoka & Hirsh, 2010; Nation, 2013). Cobb conducted a quantitative study of in-text frequency of word families with BNC frequency levels at the first three thousand levels⁷. He compiled a 517,000 token corpus of fiction, press writing and academic writing taken from the Brown corpus and searched for repetitions of ten word families from each frequency level. Cobb found that nearly all of the 1,000 level words were repeated more than six times in each corpora, but only half of the 3,000 level words were repeated six or more times (Cobb, 2007). From the findings in his study, Cobb claimed that "words beyond the 2,000 most frequent are unlikely to be encountered in natural reading in sufficient numbers for consistent learning to occur" (Cobb, 2007, p. 60). His claim was contested by McQuillan and Krashen, and there is evidence that it is possible to read enough input in order to acquire vocabulary at this level (McQuillan & Krashen, 2008; Nation, 2013). However, all agree that this is a time consuming process and requires large amounts of reading (Cobb, 2007; Krashen, 2013; Nation, 2013).

Further support is found again, in the study conducted by Matsuoka & Hirsh (2010). Their examination of AWL word families in 12 textbook texts represented in one textbook showed that over 40% of the AWL word families used in these texts occurred only once. (2010, p. 64). For the current study over 60% of the AWL vocabulary only occurred once, across three and four topic related texts.

Kang's study (2015) found that both receptive and productive vocabulary was acquired through narrow reading (see section 2.4.4). "Repeated encounters with the thematic

⁷ A more in-depth description of the studies presented here is found in section 2.4.2.

concept appeared to help learners develop semantic networks around the [target] words... Frequent encounters with target words in recurring contexts [also] helped their learning” (Kang, 2015, p. 175). The current study has only examined in-put and not tested students’ vocabulary acquisition; however, it was found that narrow reading would increase the repetition of AWL word families slightly. Usage-based theory also supports the need for exposure to words in different contexts (N. Schmitt, 2010) which may be provided through narrow reading.

Perhaps the greatest implication to be gain from the findings from this study discussed so far is that students need to be made aware of their choices when it comes to how best to acquire general academic vocabulary. Some will be able to read enough written input to acquire large quantities of vocabulary, those that cannot should be given other options.

5.3 Glossing

Several different analyses were conducted to provide answers to the research question related to glossing i.e., to what extent the use of glossaries in tailored texts assist the acquisition of general academic vocabulary. Glossary coverage in relation to percentage of text glossed was calculated, and glossary items were analyzed to find the extent of AWL word families and frequency levels related to these glossed AWL words.

5.3.1 Glossary Coverage

As shown in section 4.2.1, the total glossary coverage was below the recommended minimum of 3% (Nation, 2013). Large variations between the texts were also found, from 1.2% to 4.6% glossary coverage. This suggests that there is room for more glossary use in all of the 15 tailored texts analyzed for the current study. Words need to be noticed in order to be learned (Schmidt, 2001). Glossing is a form of awareness that can easily help learners gain better word knowledge and help them expand their vocabulary size. The use of glossaries may help L2 learners become aware of words that are not repeated in the text with only minor interruptions in reading (Nation, 2013).

Results from the analyses of AWL word families included as glossary items showed that, on average, under 25% of the glossary words in this corpus are found on the AWL. The textbook *Targets* showed the highest average rate at just under 25% (see section 4.2.2). This

could be expected as Targets also had the highest average percentage of AWL vocabulary use in total, at nearly 7% (see tables 7 and 8). On average, around 20% of the AWL terms used in the text were in the glossaries; conversely, between 70% and 80% of the AWL word families used in these tailored texts were not glossed. A majority of the AWL word families included in the glossaries occurred only once in the text and glossing of these words can aid L2 acquisition during unassisted reading (see section 4.2.2). However, a majority of the AWL word families used once in the text are not included in the glossary. The findings suggest that glossary term are chosen rather randomly. A majority of the AWL word families not defined in the glossary were from BNC/COCA 3,000 and 4,000 frequency levels. For a detailed description, see section 4.2.2.

The analyses also suggest that the recommendation to prioritize mid-frequency vocabulary could be followed more closely (Nation, 2013). Glossing mid-frequency vocabulary is important to help learners expand their vocabulary size in a relatively easy fashion. Because glossing may help learner comprehension without disrupting the reading process to any great extent they can also be seen as an important aid to the vocabulary acquisition process described in the Lexical Quality Hypothesis (see section 2.4.4). Through increased vocabulary size, L2 learners will be able to comprehend more of what they read and in doing so will also facilitate more vocabulary acquisition because they are able to read and comprehend more text.

Findings from these analyses also support claims that general academic vocabulary is not widely defined in glossaries (Flowerdew, 1993). AWL words that are at the 3,000 frequency level or above should either be glossed more or taught explicitly though such means as pre-teaching before reading the terms in a text.

Frequency is not the only essential factor involved in implicit vocabulary acquisition. The concepts of attention, awareness and noticing, are also important in a theoretical discussion of L2 vocabulary learning. As such, the use of glossary items that are in focus for the current study may both provide aid for *noticing* and lead the L2 learner towards *understanding* (Schmidt, 1995). The use of L1 translations may also help “establish an initial form-meaning link” which lead to the development of awareness at the level of understanding at once (N. Schmitt & Verspoor, 2013, p. 357).

5.4 Lexical Coverage

The main research question related to the current study has been to assess to what extent the use of general academic vocabulary in factual, textbook texts provide the means for the implicit acquisition of this vocabulary during unassisted reading. Findings from the current study have so far been discussed in relation to AWL frequency and glossary use. The final discussion takes a closer look at findings related to reading comprehension as expressed in lexical coverage rates and expected vocabulary size.

5.4.1 General Lexical Coverage

Empirical evidence shows that broad vocabulary knowledge is necessary for L2 learners to comprehend what they read (Laufer, 2010; N. Schmitt et al., 2015). In order to reach appropriate comprehension levels during unassisted reading most researchers agree that L2 learners need a lexical coverage of 98%, though 95% can be adequate. Lexical coverage refers to the “percentage of the vocabulary in a stretch of spoken or written discourse [which] needs to be known by a learner in order for him or her to understand the discourse” (N. Schmitt et al., 2015, p. 2). As such, it is closely related to vocabulary size. For the current study vocabulary size has been discussed in terms of L2 learner comprehension at the level of form-meaning, receptive word knowledge only.

To achieve an optimal lexical coverage of 98% in a majority of the 21 analyzed texts (66.7), L2 readers would need a vocabulary size up to and including the 9,000 BNC/COCA frequency level i.e., mid-frequency vocabulary. Twenty percent of the texts in the current study did not provide 98% lexical coverage even with mid-frequency vocabulary (see figure 3). The use of course materials in classroom situations often provides aid for students so that adequate understanding of texts in these situations may occur at 95% lexical coverage (Nation, 2013). A majority of the texts provided 95% lexical coverage with knowledge of high-frequency vocabulary i.e., a vocabulary size equal to the first 3,000 BNC/COCA frequency levels (see figure 4). These findings suggest that many of the texts in this corpus may be comprehended adequately with in class vocabulary support. This in class support is one way for teachers to enhance the written materials for learners by raising their awareness of AWL word families through teacher made glosses or explicit teaching techniques. As expressed in the Noticing Hypothesis, L2 acquisition demands attention. This is especially important when vocabulary is not used frequently enough in a text to promote implicit

learning. As Schmidt postulates, “[l]earning...is...largely...a side effect of attended processes” (Schmidt, 2001, p. 29).

Findings in the current study suggest that during unassisted reading, a majority of the texts may provide vocabulary challenges for advanced L2 readers, even though they contain relatively small percentages of AWL vocabulary. As explained by Perfetti and Hart “skill in reading comprehension rests to a considerable extent on knowledge of words” (C. A. Perfetti & Hart, 2002). The relationship is outlined in the Lexical Quality Hypothesis showing that an increase in vocabulary learning skills will lead to an increase in reading comprehension (Nation, 2013). That seems simple enough, but findings from the current study suggest for L2 learners to acquire vocabulary beyond the 2,000 frequency level, also AWL vocabulary, it is important for materials designers and teachers alike to help students become aware of the words they may not *notice*. When vocabulary is not repeated enough to elicit implicit learning, such as is often the case with word families at the 3,000 frequency level and more, because of Zipf’s law, *awareness at the level of noticing* is needed as a starting point for L2 learners to become *aware at the level of understanding* (Schmidt, 1995).

Perhaps one of the most surprising findings related to lexical coverage relates to differences found between authentic and tailored texts. A slightly larger number of tailored texts failed to provide 98% lexical coverage, even with mid-frequency vocabulary. A larger number of the authentic texts reached 95% lexical coverage with high-frequency level vocabulary size, a little over 80% of the authentic texts versus 60% of the tailored texts (see figures 3 and 4). These findings suggest that the authentic texts can be easier to comprehend, even though they used more academic vocabulary. Contrary to what (Hellekjær, 2012b) has feared, findings from the current study seem to suggest that the overall vocabulary use in these textbooks may be challenging enough, but students will not gain greater word knowledge from implicit acquisition simply from reading textbook texts because these will not provide enough repetitions of the vocabulary they need. However, the use of supplementary authentic texts will help facilitate more exposure to AWL vocabulary and should therefore not be forgotten in connection with classroom teaching. It must be remembered however that because of Zipf’s law, to acquire AWL vocabulary at the 3,000 frequency level and above, learners must have large amounts of exposure to acquire vocabulary implicitly, which can be a time consuming process (Cobb, 2007; Nation, 2013).

These findings also support previous research in this field. Nation’s (2006) study of general written English vocabulary in authentic novels and newspapers showed the need for a

vocabulary size at the 8,000 – 9,000 frequency levels to reach 98% lexical coverage. The lexical coverage rates for the current corpus correspond largely with his results.

The current study largely supports findings from Matsuoka & Hirsh (2010). They also examined lexical coverage and vocabulary frequency. Though the corpus analysis is conducted with the use of the GSL and AWL to determine frequency levels, and not the BNC/COCA I have used, the researchers also found that, high-frequency vocabulary (the GSL and AWL here) provided 95% lexical coverage in 75% of the texts (Matsuoka & Hirsh, 2010, p. 64), compared to 67% for the current study. None of the twelve textbook texts represented in Matsuoka & Hirsh's study produced lexical coverage of 98% with high-frequency vocabulary. Three of the 21 texts for the current study did produce 98% lexical coverage with high-frequency vocabulary, however.

The current findings indicate that even though tailored texts contain relatively few AWL word families, they contain enough mid-frequency vocabulary to challenge the vocabulary size of advanced L2 learners. Without help from noticing techniques, many of the textbook text may be difficult for learners to comprehend. If they do not comprehend enough of the vocabulary, the bond between reading comprehension and vocabulary size will be broken and vocabulary acquisition will be deterred.

Findings also suggest that the use of authentic texts in classroom situations is important to increase the amount of L2 learner exposure to AWL vocabulary, as well as mid-frequency vocabulary. However, it should be noted that, due to Zipf's law, a lot of reading must occur for students to gain implicit vocabulary knowledge of words at these frequency levels.

5.4.2 AWL Lexical Coverage

As findings from this study indicate, using general academic vocabulary in texts does not necessarily make them difficult for advanced L2 learners to comprehend because, contrary to earlier beliefs, many of the AWL word families are found within the first 2,000 frequency levels advanced L2 students can be expected know.

Analyses of AWL vocabulary into BNC/COCA frequency levels revealed differences between textbooks that are worth noting. In *Access to English*, the average mid-frequency coverage for AWL vocabulary was 8%. This is the highest level of mid-frequency AWL vocabulary use of the three textbooks. At the same time, *Access to English* also had the lowest

percentage of AWL vocabulary use in this study at an average of 4,8% (tables 3 and 4). These findings could indicate that overall text comprehension may be difficult even though the use of AWL vocabulary is at a minimum. The use of mid-frequency AWL vocabulary relates to vocabulary that may not be a part of advanced L2 learner vocabulary size. If too many mid-frequency words are used without being glossed even advanced L2 learners may struggle to comprehend the text due to the causal relationship between reading comprehension and vocabulary acquisition (Nation, 2013; C. A. Perfetti & Hart, 2002).

For *Stunt* the average of mid-frequency level vocabulary was 7.2%. On average, a majority of AWL vocabulary used in *Stunt* was from the first 2,000 frequency levels. However, these findings seemed largely due to wide diversity between texts, which suggests that L2 learner comprehension may vary greatly from text to text.

Though *Targets* showed the highest average percentage use of AWL vocabulary of the three analyzed textbooks, the use of mid-frequency level AWL was lowest. This means, for reasons explained above, that though there was a relatively high average use of AWL word families in *Targets* the AWL vocabulary used here may be more easily comprehended by L2 learners.

The current study only focuses on written input provided for students in course materials; however, it can also be productive to compare findings here to Scandinavian studies related to student production. A 2008 study of Danish 15 and 16-year-old students found that most did not have receptive, form-meaning knowledge of the first 2,000 frequency levels as defined by the GSL (see section 2.4.2). If these results are any indicator of the vocabulary size relevant to the target group for the analyzed textbooks in the current study, many students will have difficulties comprehending the texts in this corpus. Speaking as an experienced EFL teacher, what can be certain is that in every classroom students' vocabulary size will vary widely.

One interesting aspect of Stæhr's study is his decision to excluded the academic word level from the Vocabulary Levels Test (VLT) because "it is not relevant for low-level learners" (2008, p. 143). At the same time he poses, "the 2000 vocabulary level is a crucial learning goal for low-level EFL learners". Findings from the current study support previous research (Cobb, 2010) which indicates that many of the AWL vocabulary words are found at the 2,000 frequency level. With these findings in mind, it is perhaps necessary to reassess the importance of teaching general academic vocabulary as well as when to start.

Few studies have been conducted in this field in Norway. Langeland's longitudinal study of Norwegian student production was conducted with the use of vocabulary tests and computer profiling programs (see section 2.4.4). These instruments were used in order to track receptive and productive vocabulary development among nine to thirteen year old students. She found that for productive vocabulary use the students "[were greatly dependent upon] the first 1,000 words...but ... were gradually making use of a larger vocabulary" (2012, p. 140). The findings in this study also indicated an uneven development of receptive vocabulary. The rise in receptive vocabulary was "more than double, between 2008 and 2009 compared to the rise between 2009 and 2010" (Langeland, 2012, p. 135). The slow, uneven growth of vocabulary acquisition found in this study supports the assumption that the L2 English vocabulary acquisition process is demanding. It also shows that these students will need to learn a lot of vocabulary during the next few years in order to comprehend high school level texts. I would argue that a comparison of this study with my own again shows the need for greater attention to general academic vocabulary in the classroom.

Due to the causal relationship between reading comprehension and vocabulary size described in the Lexical Quality Hypothesis, the findings from this study indicate that pre-teaching of AWL vocabulary would be beneficial to advanced L2 students. In this I support claims made by Matsuoka & Hirsh (2010) following their study. They claim that textbook texts may provide "[an] opportunity...for learners to focus on academic vocabulary. This would improve reading comprehension and provide a good return for learning effort for students on an academic pathway" (2010, pp. 63-64). Pre-teaching would both enhance *awareness at the level of noticing* i.e., a conscious attention to the form of a word, but can also facilitate *awareness at the level of understanding* i.e., strengthening form-meaning knowledge (Schmidt, 1995). By translating and or define some words explicitly the L2 learners should be able to enhance their vocabulary size which in turn would strengthen reading comprehension to further assist the process of strengthening vocabulary size through increased exposure to usage events (C. A. Perfetti & Hart, 2002; N. Schmitt & Verspoor, 2013).

5.5 Brief Summary of Findings

The key findings from the current study show that a large majority of the AWL word families used in the corpus lack adequate exposure and repetition for the implicit acquisition of these AWL families. A majority of the word families repeated enough to promote implicit

acquisition were within the first 2,000 frequency levels, and may be words already within the students' vocabulary size.

Many of the AWL word families present in the corpus were found within the first 2,000 frequency levels in which it may be expected that advanced L2 learners have form-meaning receptive knowledge of. If, however, prior teaching practices and course materials have not provided students with repeated exposure to or noticing of these word families, usage-based theory dictates that they will not have been acquired (Langacker, 2000; N. Schmitt & Verspoor, 2013).

Frequency of repetition is not the only aspect of learning that promotes the vocabulary acquisition process. Implicit vocabulary acquisition is also at least partly dependent on enhancing L2 learners attention to the words being acquired (Schmidt, 2001; N. Schmitt & Verspoor, 2013). If noticing is not facilitated by exposure and frequency, something the findings here indicate with regard to AWL usage, then glossing may be an effective means of raising student awareness to AWL vocabulary (R. Ellis & Shintani, 2014). Findings from the current study show that a majority of the AWL word families were not glossed. Though glossing is an effective means of helping learners comprehend texts with a lexical coverage just outside of their vocabulary size not all words can be glossed. Recommended glossary coverage rates are between 3%-5% (Nation, 2013). Findings related to glossary coverage indicate that all of the tailored text could make more extensive use of glossaries. By glossing AWL word families, textbook designers and teachers may help L2 learners acquire these words implicitly by raising learner awareness. Also providing L1 translations can help learners gain better word knowledge more quickly as they may have abstracted the terms already in the L1 and can therefore related this meaning to the L2. It should also be pointed out that such transfer is not always appropriate because different cultural meaning, among other things, may color the way a word is used (Nation, 2013; N. Schmitt & Verspoor, 2013).

Linguists agree that vocabulary should be taught in context. Usage-based theory defines the acquisition of form-meaning understanding through cognitive processes involving associations related to reoccurring instances of symbolic units within different contexts (Langacker, 2000; N. Schmitt & Verspoor, 2013). This does not, however, exclude the need for awareness, both at the level of noticing and understanding. Findings here suggest that few AWL vocabulary words will be learned implicitly. Explicit teaching of some relevant AWL vocabulary words before reading texts may then be necessary in order to aid students in the awareness process that can lead them to acquisition of these terms.

6. Conclusion

The research questions guiding this corpus based study are directly related to the implicit acquisition of general academic vocabulary through unassisted reading.

1. To what extent does the use of general academic vocabulary in factual, textbook texts provide the means for the implicit acquisition of this vocabulary during unassisted reading?

1a. How is general academic vocabulary used within factual, textbook texts and across topic related texts?

1b. To what extent does the use of glossaries in tailored texts assist advanced L2 English learners with the acquisition of general academic vocabulary during unassisted reading?

In order to gain answers to these research questions, a corpus study of written textbook texts was conducted. The study included mapping vocabulary use and investigating how general academic vocabulary usage in the corpus may or may not promote implicit vocabulary acquisition of AWL word families. General academic vocabulary may be defined as vocabulary common to many different academic disciplines and has been operationalized in the current study through use of Coxhead's Academic Word List (AWL). The hope has been that the study may generate new knowledge about general academic vocabulary use in course materials, and at the same time place this field of research within a Norwegian context. The course materials studied were written for first year high school students, 15- and 16-year-olds, enrolled in the last obligatory English course before qualification for university level studies in the Norwegian school system.

The corpus compiled for this study has been comprised of 21 factual, textbook texts in three different textbooks, containing a total of 28,734 tokens. The investigation of the corpus was conducted with the use of the VocabProfiler's VP-classic and VP-Compleat programs as well as the Range program, all of which were found on the Lextutor website (Cobb, n.d.-b).

In the following chapter, key findings will be presented in relation to the research questions. Other concluding remarks will encompass contributions of the study, possible implications of findings and recommendations for further research in this area of enquiry.

6.1 Key Findings

Key findings from the current study show a number of tendencies in relation to the use of AWL vocabulary and implicit acquisition through reading factual, textbook texts.

6.1.1 AWL Usage

The present study showed greater AWL word family coverage in authentic texts than tailored texts, on average 7.5% versus 4.1%. A slight majority of the texts in the corpus provided an AWL coverage rate of 5% or more. Even so, over 45% of the texts showed coverage rates below 5%. By using both factual and authentic texts, AWL coverage rates were around rates expected in English language newspapers, ca 5%. Despite being factual texts, the majority of texts in this corpus (85%) cannot be compared to academic texts i.e., which normally contain between 8%-10% AWL coverage (Coxhead, 2000). These findings suggest that it is important to also include authentic written texts in classroom reading.

A large majority of the AWL word families used in the corpus, over 60% across topic related texts, fail to provide adequate exposure and repetition for the implicit acquisition of these AWL families. Under 3% of the AWL word families in the corpus were recycled the minimum recommended frequency of six or more times. A majority of the word families in this category were within the first 2,000 frequency levels and may be words students know. Sixty percent of the AWL word families used across topic related texts were only used once. These findings indicate that, given low rates of recycling, it is unlikely for a majority of the AWL word families to be acquired implicitly through unassisted reading, implying that explicit means of teaching AWL word families should strongly be considered.

6.1.2 Glossing

The implicit acquisition of AWL vocabulary was also investigated in relation to the use of glossaries to enhance L2 learner awareness. Recommended glossary coverage rates are between 3%-5% (Nation, 2013). Findings from the current study show that all of the tailored texts could make more extensive use of glossaries. The average rate of total glossary coverage for the 15 tailored texts that included glossaries was 2.7%. Because glossing can be an efficient way for L2 learners to acquire vocabulary material designers could make wider use of glossaries.

A majority of the AWL word families included in the glossaries occurred only once in the text so that glossing can aid L2 acquisition of these terms during unassisted reading. Findings also showed that a majority of the AWL word families in this corpus, between 70% and 80% per text, were not glossed and a majority of these were from BNC/COCA 3,000 and 4,000 frequency levels. This indicates that there are many AWL word families perhaps outside the vocabulary size of advanced L2 learners that are not recycle enough to promote implicit vocabulary acquisition and are also not glossed. These findings strongly support the need for explicit attention to AWL word families if advanced L2 learners are to acquire them.

6.1.3 Lexical Coverage

To achieve an optimal lexical coverage of 98% in a majority of the 21 analyzed texts L2 readers would need a vocabulary size up to and including the 9,000 BNC/COCA frequency level i.e., mid-frequency vocabulary. A majority of the texts (66.7%) provided 95% lexical coverage with knowledge of high-frequency vocabulary i.e., a vocabulary size equal to the first 3,000 BNC/COCA frequency levels. These findings suggest that many of the texts in this corpus may be comprehended adequately with in-class vocabulary support, but unassisted reading may be difficult for many.

There were surprising differences between authentic and tailored texts in relation to lexical coverage. While lexical coverage at 98% is very similar at high and mid-frequency levels, more tailored text do not reach 98% coverage with mid-frequency vocabulary, 20% versus 16.7% for authentic texts. At the same time, a larger number of the authentic texts reached 95% lexical coverage with high-frequency level vocabulary size, 83% versus 67% respectively. These findings suggest that the authentic texts can be easier to comprehend, even though they used more academic vocabulary, supporting the need to use authentic texts in classroom situations.

Finally, the general findings related to AWL vocabulary all confirm that a large number of AWL word families fall within the first 2,000 frequency levels. With such high rates of frequency in authentic corpora, this finding points towards the importance of general academic vocabulary as a learning goal for L2 learners also in Norway.

6.2 Contributions

It is my hope that the present study can contribute to a better understanding general academic vocabulary usage in English subject course materials and how this usage may effect L2 vocabulary acquisition. I hope my findings contribute to a better understanding of vocabulary use in factual, textbook texts written for advanced L2 learners of English and that it may help others continue to use corpus linguistics as a means of gaining more knowledge in this field of study. I also hope the study may contribute to greater insight into the study of implicit vocabulary acquisition through reading, from the viewpoint of text analyses rather than learner production. I also hope that the present study will contribute to placing the discussion of general academic vocabulary acquisition into the Norwegian educational context. The study can possibly also contribute to a better understanding of how different computer analyses can work together to provide more in-depth analyses that help enlighten the study of vocabulary acquisition in general.

6.3 Implications

6.3.1 Materials design

It is clear that textbook authors must place their main focus on content when writing factual texts and repeating words incessantly can decrease the coherence of any text. As such, it is not the aim of this study to encourage these authors to write differently. I would like to think, however, that they might gain insight into ways in which computer programs can aid their understanding of the vocabulary choices they make. This means that they should be encouraged to use general academic vocabulary in texts and otherwise when selecting and producing course materials. The findings from the current study clearly show the advantages of authentic texts in relation to AWL exposure. As such, materials designers should be encouraged to use authentic, factual texts, to an even greater degree than is the case today, both on websites and in textbooks.

Computer programs such as those used for the current study (luxtutor.ca), together with recognition for recommendations made by researchers would perhaps be even more helpful for those designing glossaries. Using higher levels of glossary coverage and glossing mid-frequency vocabulary to an even greater extent would most likely help L2 learners greatly.

6.3.2 Classroom practices

In the classroom environment, time is of the essence. Perhaps instead of discussing many topics briefly, teachers in Norway may consider spending more time on fewer topics, since the curriculum provides some opportunity to form topic choices. In this manner, students could profit from the benefits of narrow reading and a little more time could be spent on setting academic vocabulary goals for students and helping them acquire the knowledge they need to expand their vocabulary.

I strongly encourage teachers and material designers alike to place vocabulary acquisition in focus. As such, I hope the current study may help provide input that can help teaching practices and material design.

It also seems important that vocabulary acquisition theory and teaching practices receives the attention it deserves during teacher training.

6.4 Recommendations for Further Study

There are many areas in which this study may be used as a starting point for further research. A more in-depth study of the differences between tailored, factual texts and authentic, factual texts linked to English course materials may provide even better insight into course material uses in the Norwegian education system and would be of benefit to both teachers and material designers. The current findings also suggest that future studies in the field may consider placing AWL word families into BNC/COCA frequency levels, at least until revisions are made or further studies conclude with other ways of analyzing general academic vocabulary as this may provide more in-depth information during corpus analyses. The findings also seem to indicate that further study of glossary use in course materials could provide better understanding in the field and help develop even better course materials. Follow-up studies similar to the one conducted here could be applied to textbooks for elective English courses in the second and third years of high school studies. A study of this type would be able to assess if AWL vocabulary is used more extensively in these textbooks. Finally, it would also seem that the current study shows the need for further research into productive and receptive vocabulary size among L2 learners of English in Norway. Little is known about Norwegian learners vocabulary and not only would a study of this kind build upon knowledge gained in the current study, it could also complement Hellekjær's research on reading comprehension and reading strategies (see section 1.3). A study of this type could also help researchers better

understand the implications of the current study and may help material designers, teachers and students alike improve vocabulary learning practices.

References

- Aarts, B. (2000). Corpus Linguistics, Chomsky and Fuzzy Tree Fragments. In C. Mair & M. Hundt (Eds.), *Corpus Linguistics and Linguistic Theory* (pp. 5-14). Amsterdam: Editions Rodopi B.V.
- Areklett, K. M., Hals, Ø., Lindaas, K., & Tørnby, H. (2009). *Stunt*. Bergen: Fagbokforlaget.
- Balsvik, L., Bratberg, Ø., Henry, J. S., Kagge, J., & Pihlstrøm, R. (2015). *Targets*. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Baumann, J. F., & Graves, M. F. (2010). What Is Academic Vocabulary? *Journal of Adolescent & Adult Literacy*, 54(1), 4-12. doi:10.2307/20749071
- Bonelli, E. T. (2012). The Evolution of Corpus Linguistics. In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14-28). Abingdon, Oxon: Routledge
- Brown, C. (2015). Word Knowledge. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Learning Acquisition* (pp. vi-xxiv, 1-756). New York: Routledge.
- Brown, C., Cullican, B., & Phillips, J. (2013). A New General Service List. Retrieved from <http://www.newgeneralservicelist.org/>
- Burgess, R., & Sørhus, T. B. (2013). *Access to English*. Oslo: Cappelen Damm AS.
- Charboneau, R. (2012). Approaches and Practices Relating to the Teaching of EFL Reading at the Norwegian Primary Level. In A. Hasselgren, I. Drew, & B. Sørheim (Eds.), *The Young Language Learner* (pp. 51-69). Bergen: Fagbokforlaget.
- Cobb, T. (2007). Computing the Vocabulary Demands of L2 Reading. *Language Learning & Technology*, 11(3), 38-63. Retrieved from <http://lt.msu.edu/vol11num3/cobb/>
- Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, 22(1), 181-200. Retrieved from <http://nflrc.hawaii.edu/rfl>
- Cobb, T. (2015). Web Vocabprofile. Retrieved from <http://www.lextutor.ca/vp/>
- Cobb, T. (n.d.-a). The original idea behind this website Why & how to use frequency lists to learn words. Retrieved from <http://www.lextutor.ca/research/rationale.htm>
- Cobb, T. (n.d.-b). Web VP Classic v.4 Retrieved from <http://www.lextutor.ca/vp/eng/>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6 th ed.). Oxon: Routledge.
- Corson, D. (1997). The Learning and Use of Academic English Words. *Language Learning*, 47(4), 671-718.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. (2006). *Essentials of Teaching Academic Vocabulary*. Boston: Heinle, Centage Learning.
- Coxhead, A. (2011). The Academic Word List ten years on: Research and teaching implication. *TESOL Quarterly* 45(2), 355 - 362.
- Coxhead, A. (2012). What can corpora tell us about English for Academic Purposes? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. ix-xxvii, 1-682). New York: Routledge.
- Coxhead, A. (n.d.). Sublists of the Academic Word List. Retrieved from <http://www.victoria.ac.nz/lals/resources/academicwordlist/publications/awlsublists1.pdf>
- Creswell, J. W. (2014). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, Mass.: Pearson.
- Dörnyei, Z. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

- Elgort, I. I., & Warren, P. P. (2014). L2 Vocabulary Learning From Reading: Explicit and Tacit Lexical Knowledge and the Role of Learner and Item Variables. *Language Learning*, 64(2), 365-414.
- Ellis, N. C. (1994). *Implicit and explicit learning of languages* (N. C. Ellis Ed.). London: Academic Press.
- Ellis, N. C. (2002). Reflections on Frequency Effects in Language Processing. *Studies in Second Language Acquisition*, 24(2), 297-339.
- Ellis, N. C. (2012a). Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics*, 32, 17-44.
- Ellis, N. C. (2012b). Frequency-based Accounts of Second Language Acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 193-210). New York, NY: Routledge.
- Ellis, N. C. (2015). Frequency Effects. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. vi-xxiv, 1-755). New York: Routledge.
- Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. Oxon: Routledge.
- Flowerdew, J. J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231-244.
- Gardner, D., & Davies, M. (2013). A New Academic Vocabulary List. *Applied Linguistics* 2014, 35(3), 305-327. Retrieved from <http://applied.oxfordjournals.org/>
- Gilmore, A. A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97-118.
- Grabe, W. W. (2008). *Reading in a Second Language Moving from Theory to Practice* (1 ed.). Cambridge: Cambridge University Press.
- Gulden, A. T. (2008). English for Academic Purposes: A New Discipline in Norway? *Nordic Journal of English Studies*, 7(3), 207-211.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4 (2), 237-258.
- Heibert, E. H., & Lubliner, S. (2008). The Nature, Learning, and Instruction of General Academic Vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What Research Has To Say about Vocabulary Instruction* (pp. 106-129). Newark, DE: International Reading Association.
- Hellekjær, G. O. (2005). *The acid test: does upper secondary EFL instruction effectively prepare Norwegian students for the reading of English textbooks at colleges and universities?* (Vol. nr 240). Oslo: Det humanistiske fakultet, Universitetet i Oslo.
- Hellekjær, G. O. (2008). A Case for Improved Reading Instruction for Academic English Reading Proficiency. *Acta Didactica Norge*, 2(1), 1-17.
- Hellekjær, G. O. (2009). Academic English Reading Proficiency at the University Level: A Norwegian Case Study. *Reading in a Foreign Language*, 21(2), 198-222. Retrieved from <http://nflrc.hawaii.edu/rfl>
- Hellekjær, G. O. (2012a). Engelsk - faget som ikke utfordrer elevene i den videregående skolen. *Communicare*, 2, 30-33.
- Hellekjær, G. O. (2012b). Fra Reform 94 til Kunnskapsløftet: en sammenligning av leseferdigheter på engelsk blant avgangselever i den videregående skole i 2002 og 2011 *Kvalitet i norsk skole* (pp. 153-170). Oslo: Kvalitet i norsk skole: internasjonale og nasjonale undersøkelser av læringsutbytte og undervisning.
- Hestetræet, T. (2012). Teacher Cognition and the Teaching and Learning of EFL vocabulary. In A. Hasselgren, I. Drew, & B. Sørheim (Eds.), *The Young Language Learners* (pp. 177-190). Bergen: Fagbokforlaget.

- Hyland, K. (2011). Disciplinary Specificity: Discourse, Context, and ESP. In D. Belcher, A. M. Johns, & B. Paltridge (Eds.), *New Directions in English for Specific Purposes Research* (pp. iii-v, 1-282). Michigan: The University of Michigan Press.
- Hyland, K., & Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, 41(2), 235-253. doi:10.2307/40264352
- Index, E. F. E. P. (n.d.). English First English Proficiency Index. Retrieved from <http://www.ef.no/epi/>
- Juuhl, G. K., Hontvedt, M., & Skjelbred, D. (2010). *Læremiddelforskning etter LK06 Eit kunnskapsoversyn*. Retrieved from Tønsberg:
- Kang, E. Y. E. (2015). Promoting L2 Vocabulary Learning through Narrow Reading. *RELC journal*, 46(2), 165-179.
- Kemmer, S., & Barlow, M. (2000). *Usage-based models of language*. Stanford, Calif: CSLI Publications.
- Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press Ltd.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Krashen, S. D. (2013). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research*, 1(1), 27.
- Langacker, R. W. (2000). A Dynamic Usage-Based Model. In M. Barlow & S. Kemmer (Eds.), *Usage Based Models of Language* (pp. 1-60). Standford: Center for the Study of Language and Information.
- Langeland, A. S. (2012). Investigating vocabulary development in English from grade 5 to 7 in a Norwegian primary school. In A. Hasselgren, I. Drew, & B. Sørheim (Eds.), *The Young Language Learner* (pp. 131-143). Bergen: Fagbokforlaget.
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.
- Laufer, B. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Leow, R. P. (2013). Attention in SLA. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 41-44). New York: Routledge.
- Lesaux, N. K., Keiffer, M. J., Kelley, J. G., & Harris, J. R. (2014). Effects of Academic Vocabulary Instruction for Linguistically Diverse Adolescents: Evidence From a Randomized Field Trial. *American Educational Research Journal*, 1-36. Retrieved from <http://aer.sagepub.com/content/early/2014/04/25/0002831214532165>
- Mahan, K. R., & Brevik, L. M. (2013). «I can English very good» – engelske ordfeil blant norske elever og studenter. *Bedre Skole*, 3, 34-39.
- Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22(1), 56-70.
- McCarthy, M., & O'Keeffe, A. (2012). Historical Perspective. In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. v-xxvii, 1-682). Abingdon, Oxon: Routledge.
- McQuillan, J., & Krashen, S. S. (2008). Commentary: can free reading take you all the way? A response to Cobb (2007). *Language Learning & Technology*, 12(1), 104.
- Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly*, 47(1), 91-108. doi:10.1002/RRQ.011
- Nation, I. S. P. (2006). How Large a Vocabulary Is Needed For Reading and Listening? *The Canadian Modern Language Review*, 63(1), 59-82.

- Nation, I. S. P. (2012). *The BNC/COCA word families lists*. Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (second ed.). Cambridge: University Printing House.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13. Retrieved from <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Vocabulary-Size-Test-information-and-specifications.pdf>
- Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Sherrise Roehr.
- Nation, I. S. P., & Webb, S. S. (2008). *Evaluating the vocabulary load of written text*. <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/2008-Webb-Evaluating-vocabulary-load.pdf>.
- Park, D. (n.d.). Identifying & Using Formal & Informal Vocabulary. Retrieved from http://www.academia.edu/8186664/Identifying_formal_and_informal_vocabulary
- Pellicer-Sánchez, A. (2015). INCIDENTAL L2 VOCABULARY ACQUISITION FROM AND WHILE READING. *Studies in Second Language Acquisition*, 1-34.
- Perfetti, C. (2010). Decoding, Vocabulary and Comprehension In M. G. McKoewn & L. Kucan (Eds.), *Bringing Reading Research to Life* (pp. 291-303). New York: The Guilford Press.
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of Functional Literacy* (pp. 189-213). Amsterdam: John Benjamins Publishing Company.
- Perfetti, C. C., & Stafura, J. J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific studies of reading*, 18(1), 22-37.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. New York: Longman Inc.
- Read, J. (2000). *Assessing Vocabulary*. New York: Cambridge University Press.
- Richards, J. C., Schmidt, R. W., & Richards, J. C. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed. ed.). London: Longman.
- Rott, S. S. (2002). The Effect of Multiple-Choice L1 Glosses and Input-Output Cycles on Lexical Acquisition and Retention. *Language Teaching Research*, 6(3), 183-222.
- Ruegg, R., & Brown, C. (2014). Analyzing the Effectiveness of Textbooks for Vocabulary Retention. *Vocabulary Learning and Instruction*, 3(2), 11-18. Retrieved from <http://vli-journal.org>
- Schmidt, R. (1995). Consciousness and Foreign Language Learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning* (pp. 1-63). Honolulu: University of Hawaii.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.
- Schmidt, R. (2010). *Attention, Awareness, and Individual Differences in Language Learning*. Paper presented at the CLaSIC 2010, Singapore.
- Schmitt, D., & Schmitt, N. (2012). Plenary Speech A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Cambridge Journal*, 1-20. Retrieved from <http://journals.cambridge.org>
- Schmitt, N. (2008). Review Article: Instructed Second Language Vocabulary Learning. *Language Teaching Research*, 12(3), 329-363. doi:10.1177/1362168808089921
- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Hampshire: Palgrave Macmillan.

- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2015). How much vocabulary is needed to use English? Replication of vanZeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 1-15. Retrieved from http://journals.cambridge.org/abstract_S0261444815000075
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1), 26-43. doi:10.2307/41262309
- Schmitt, N., & Verspoor, M. (2013). Language and the Lexicon in SLA. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 353-360): New York/London: Routledge.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), 487-512. doi:10.1093/applin/amp058
- Stella, S. C. (2015). *Creating an Academic Business English List: a Corpus Based Study*. (Degree of Masters of Arts (Applied Linguistics) MA), Concordia University, Montreal.
- Stæhr, L. S. L. (2008). Vocabulary size and the skills of listening, reading and writing. *Language learning journal*, 36(2), 139-152.
- Taylor, C. (2008). What is corpus linguistics? What the data says. *International Computer Archive of Modern and Medieval English Journal*, 32, 179-200.
- Utdanningsdirektoratet. (2006). *Core curricula*. Retrieved from http://www.udir.no/Upload/larerplaner/generell_del/5/Core_Curriculum_English.pdf?epslanguage=no.
- Utdanningsdirektoratet. (2013). *Læreplan i engelsk*. Retrieved from http://www.udir.no/kl06/ENG1-03/Hele/Komplett_visning/?print=1.
- West, M. (1953). *A General service list of English words : with semantic frequencies and a supplementary word-list for the writing of popular science and technology* (Rev. and enl. ed. ed.). London: Longman.

7. Appendices

7.1 Textbook survey

7.1.1 Information to the schools

Jeg er i prosessen av å samle inn data for en erfaringsbasert master i undervisning med fordypning i engelsk, ved Universitetet i Bergen. I denne forbindelsen ønsker jeg å først få en oversikt over hvilke læreverker er i bruk for engelskfaget, vg1 studiespesialiserende. Dette for at min forskning vil kunne være mest mulig relevant til vår skolehverdag.

Jeg skal undersøke bruken av akademisk vokabular i noen av disse læreverkene. Det finnes forskning som tyder på at innlæring av akademisk vokabular før påbegynt høgskoleutdanning kan ha betydning for i hvor stor grad elevene vår lykkes med deres videre studier på universitets- og høyskolenivå.

Dere har kanskje allerede fått en henvendelse, og jeg beklager maset, men samtidig håper jeg at dere kan ta tid i en travel hverdag for å svare på denne mailen for meg. Du kan skrive navn på skolen, hvilket fylke skolen tilhører og sette kryss i vedlagt tabell. Disse opplysningene kan du bare sette inn i selve retur e-posten, jeg trenger ikke informasjonen i et Word dokument.

Mvh,

Kimberly Skjelde

Knarvik videregående skole, Hordaland Fylke

E-post: kimskj3@hfk.no

Fylke:

Navn på skolen:

Læreverket	Sett kryss
<i>Access to English</i> (Cappelen Damm, 2013)	
<i>Passage</i> (Cappelen Damm, 2009)	
<i>Passage</i> (Cappelen Damm, 2006)	
<i>Targets</i> (Aschehoug, 2009)	
<i>Gateways, SF</i> (Aschehoug, 2011)	
<i>Stunt</i> (Fagbokforlaget, 2009)	
@cross (Cappelen Damm, 2006)	
Ndla.no	
Annet	

7.1.2 Overview of replies from skoles

<i>Access to English</i> (Cappelen Damm, 2013)	Sogn og Fjordane Fylke: Firda vgs Hordaland Fylke: Tertnes videregående skole, Langhaugen (5), Knarvik vgs, Buskerud Fylke: Numedal videregående skole Oppland Fylke: Vinstra vidaregåande skule, Gjøvik videregående skole Aust-Agder Fylke: Møglestu videregående skole
<i>Passage</i> (Cappelen Damm, 2009)	Troms Fylke: Breivang Nord-Trøndelag Fylke: Grong vgs. Oppland Fylke: Gjøvik videregående skole Hordaland Fylke:: Sotra vgs, Voss gymnas
<i>Passage</i> (Cappelen Damm, 2006)	Aust-Agder Fylke: Sam Eyde videregående skole Nord-Trøndelag Fylke: Inderøy vgs
<i>Targets</i> (Aschehoug, 2009)	Aust-Agder fylke: Tvedestrand og Åmli videregående skole Buskerud Fylke: Ringerike videregående skole, Gol vgs Hordaland Fylke: Olsvikåsen vgs, Askøy videregående skole, Kvinnherad vidaregåande skule, Langhaugen (1), Osterøy vgs, Stord vidaregåande skule, Laksevåg vgs Oppland Fylke: Lena videregående skole Sogn og Fjordane Fylke: Høyanger vg skule, Sogndal vgs Dahlske vgs i Grimstad Vest- Agder Fylke: Tangen Videregående Skole Troms Fylke: Kvaløya vgs Rogaland Fylke: Jåtte vgs
<i>Gateways, SF</i> (Aschehoug, 2011)	Aust-Agder Fylke: Møglestu videregående skole Buskerud Fylke: Numedal videregående skole Oppland Fylke: Lillehammer vgs Hordaland Fylke: Stord vgs, Odda vidaregåande skule
<i>Stunt</i> (2009)	Nord-Trøndelag fylkeskommune: Levanger Vgs, Verdal videregående skole Hordaland Fylke: Øystese gymnas Aust-Agder Fylke: Risør videregående skole
@cross (2006)	
Ndla.no	Hordaland Fylkeskommune: Os vidaregåande skule, Kvinnherad vidaregåande skule Oppland Fylke: Gjøvik videregående skole Hordaland fylke: Voss gymnas
Annet	Fylke: Hordaland Fylkeskommune, Os vidaregåande skule
New Experience (Gyldendal Undervisning 2009)	Finmark fylke: Nordkapp maritime fagskole og videregående Troms Fylke: Nord-Troms vgs Vestfold fylke: Horten vgs Hordaland fylke: St. Paul gymnas

7.2 Text Analyses entire text

Each text analysis contains the following: a copy of the text only file used to produce the text analyses in the computer programs, a record of text changes made to the text only file, the text analysis from the VP-classic and VP-complete analyses, with glossary items highlighted.

7.2.1. Access to English

7.2.1.1. Divided by a Common Language

7.2.1.1.1 Text only file

POINTS OF DEPARTURE

How can you tell whether an English speaker is from Britain or the USA ? Discuss in class .

Divided by a Common Language

New nation - new language ?

It is not easy to say exactly when American English became distinct from British English . In 1776 colonists in America declared independence and started a revolution which eventually succeeded in throwing out the British . Of course , it was not easy in those days to say who was British and who was American . Many of the colonists had only been there for a generation or two . As for their speech , you would have had difficulty hearing who was a British soldier and who was a colonist . It was simply a question of deciding which side you were on : the side of revolution or the side of loyalty to the crown .

As the new nation was born , many Americans wondered what the future would bring , not only politically and economically , but also for their mother tongue . Some believed that the break with Mother England would eventually lead to the birth of a new language - American - as different from English as Norwegian is from German . Others were more impatient and actually suggested that Hebrew or Greek should be adopted as the official language of the new republic .

Two and a half centuries later two things are clear . Firstly , most Americans do not speak Hebrew or Greek ! Secondly , American and British English are still so close that it would be silly to call them different languages . In speech they are certainly much closer than many Norwegian dialects are to each other , and in writing they are much closer than the two official written norms of Nynorsk and Bokmal .

Vocabulary differences

On the other hand , if you were to eavesdrop on a conversation between an American President and a British Prime Minister you would probably be able to hear who was who . A native speaker of English certainly would . Partly it would be a question of the vocabulary they used . If one of them said , for example , " England sure looks pretty in the fall " , then you could be fairly certain that it was the President talking . The Prime Minister would be more likely to say " England certainly looks pretty in the 1 Marquess Charles Cornwallis . the British general . surrenders to American troops at Yorktown ending the War of American

Independence . Inset is a portrait of George Washington . the first president of the United States of America

autumn " . If one of them offered the other a " biscuit " , then you can be fairly certain that the Prime Minister is the host . Americans usually call them " cookies " .

There are lots of these vocabulary differences , but few of them cause any problems . British people particularly are so used to hearing American English that they are not aware of the differences . If there are misunderstandings , they are usually caused by words that have a different meaning on each side of the Atlantic . For example , " suspenders " - which in Britain are worn by ladies to keep their stockings up , and in America are worn by men to keep their trousers up . And when I say " trousers " , I mean " pants " in American English , not " pants " as in British English , which are called " underpants " in America and worn under your trousers . Unless you happen to be Superman .

Some vocabulary differences :

British English

lift lorry

pavement

American English

elevator truck sidewalk

flat apartment

underground subway

autumn fall

biscuit cookie

petrol gas

holiday vacation

shop store

rubbish garbage

sweets candy

Pronunciation differences

But even if the President and the Prime Minister avoided vocabulary differences , we would still be able to hear who was speaking by the way they pronounced their words . If one of them said " I cannot think of a better way to fight terror " , and we heard the underlined words pronounced " krent " , " beddeR " and " terreR " , then it is probably the President speaking . Americans tend to pronounce where most Britons leave them silent - " terreR " " beddeR " rather than British English " terruh " " bettuh " . Notice too how often becomes in American pronunciation . " krent " can be found in Britain too , but standard pronunciation is " kaant " .

Spelling differences

Spelling is another area where American and British English differ . Imagine that instead of eavesdropping on the conversation between these two world leaders , you were able to hack into their net chat . Would you be able to see who had written what ? Well , again it would depend on the words they chose . Early on in the history of the US it was decided that American spelling should be made more " logical " . These are some of the changes that were made - and that are still features of American English .

British English

colour , flavour , labour ... theatre , centre ... catalogue

travel - travelled - travelling plough

cheque defence

American English color , flavor , labor ... theater , center ... catalog

travel - traveled - traveling plow

check defense

Americans were very proud of their own English , and in 1820 a witty proposal was made in Congress that young English aristocrats should be invited to America so that Americans could teach them to speak properly ! It is this gentle rivalry that led the Irish writer George Bernard Shaw to describe Britain and America as two nations " divided by a common language " . Because Americans and Britons can communicate so easily , they quickly discover the cultural differences that exist between them .

So which form of English should you choose as a learner of English , British or American English ? Well , it does not really matter . Both forms are equally correct , although it is a good idea to choose one or the other rather than to mix them up . (Having said that , we should remember that there are other forms of English , e.g. Australian English and Canadian English , which combine British and American elements .)

Warning !

Finally , a word of warning . Some Norwegians imagine that American English is less " formal " than British English and that forms like " wanna " " gotta " and " ain't " are allowed . In fact these forms

can be found in both British and American English speech , but

they are not part of the standard written language in either country . There is nothing " wrong " with forms like these in themselves ,

if you use them in the right context , for example in text messages

and pop lyrics . But they do not belong in essays . It is a bit like wearing shorts - there is nothing wrong in shorts , but you would not

wear them to a funeral .

Chapter 2 GLOBALLY SPEAKING 83

7.2.1.1.2 Text changes

Divided by a Common Language

1. All glossary terms in the margin have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

wasn't, can't, it's (3), doesn't, there's (2), don't, wouldn't

3. Hyphenated words with hyphen removed:

English-speaker

4. Compound words separated:

5. Words (groups of letters) removed from the text analysis:

«tt», «dd», «r's»

6. Proper nouns list:

American, America, Australian, Canadian, Irish, George, Bernard, Shaw, Greek, Britain, English, Britons, British, Hebrew, USA, Congress, Norwegian, England, German, Atlantic, Charles, Cornwallis, Yorktown, Washington, Americans, Norwegians, superman,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Mother (England), Marquess (Charles Cornwallis), War of (American) Independence, United States of (America), nynorsk, bokmal, Prime Minister, President

Note: Texts related to illustrations have been included in the text analysis.

7.2.1.1.3 Text Analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Divided by a Common Language (6.51 kb)

Words recategorized by user as 1k items (proper nouns etc): AMERICAN, AMERICA, AUSTRALIAN, CANADIAN, IRISH, GEORGE, BERNARD, SHAW, GREEK, BRITAIN, ENGLISH, NYNORSK, BRITONS, BRITISH, HEBREW, USA, CONGRESS, NORWEGIAN, ENGLAND, GERMAN, ATLANTIC, PRIME, MINISTER, PRESIDENT, TERRUH, BETTUH, KAANT, KRENT, BEDDER, TERRER, CHARLES, CORNWALLIS, YORKTOWN, WASHINGTON, AMERICANS, NORWEGIANS, SUPERMAN, BOKMAL (total 132 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	212	268	836	78.57%
Function:	(501)	(47.09%)
Content:	(335)	(31.48%)
> Anglo-Sax	(189)	(17.76%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	40	42	50	4.70%
> Anglo-Sax:	(18)	(1.69%)
1k+2k		(83.27%)
AWL Words (academic):	18	19	24	2.26%
> Anglo-Sax:	(6)	(0.56%)
Off-List Words:	?	85	154	14.47%
	270+?	414	1064	100%

Current profile	
%	Cumul.
78.57	78.57
4.70	83.27
2.26	85.53
14.47	100.00

Words in text (tokens):	1064
Different words (types):	414
Type-token ratio:	0.39
Tokens per type:	2.57
Lex density (content words/total)	0.53

Pertaining to onlist only

Tokens:	910
Types:	329
Families:	270
Tokens per family:	3.37
Types per family:	1.22
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

A. AWL Tokens list

AWL [18:19:24] area aware communicate context cultural distinct economically elements eventually eventually features finally generation labor labour logical norms prime prime prime prime revolution revolution text

Sublist 1

area context economically labor labour

Sublist 2

cultural distinct elements features finally text

Sublist 4

communicate

Sublist 5

aware generation logical prime prime prime prime

Sublist 8

eventually eventually

Sublist 9

norms revolution revolution

B. AWL Types list

AWL types: [18:19:24] area_[1] aware_[1] communicate_[1] context_[1] cultural_[1] distinct_[1] economically_[1] elements_[1] eventually_[2] features_[1] finally_[1] generation_[1] labor_[1] labour_[1] logical_[1] norms_[1] prime_[4] revolution_[2] text_[1]

C. Families list

AWL families: [18:19:24]

area_[1] aware_[1] communicate_[1] context_[1] culture_[1] distinct_[1] economy_[1] element_[1] eventual_[2] feature_[1] final_[1] generation_[1] labour_[2] logic_[1] norm_[1] prime_[4] revolution_[2] text_[1]

AWL Fr non-cognate families: [families 3 : tokens 6] aware_[1] feature_[1] prime_[4]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, propers, cognates, extraction, barchart)

Freq. Level Families (%) Types (%) Tokens (%) Cumul. token %

K-1 Words :	233 (68.53)	283 (67.54)	887 (82.74)	82.74
K-2 Words :	49 (14.41)	55 (13.13)	86 (8.02)	90.76
K-3 Words :	32 (9.41)	35 (8.35)	38 (3.54)	94.30
K-4 Words :	7 (2.06)	7 (1.67)	9 (0.84)	95.14
K-5 Words :	9 (2.65)	10 (2.39)	16 (1.49)	96.63
K-6 Words :	3 (0.88)	3 (0.72)	3 (0.28)	96.91
K-7 Words :	3 (0.88)	4 (0.95)	4 (0.37)	97.28
K-8 Words :	1 (0.29)	1 (0.24)	1 (0.09)	97.37
K-9 Words :	1 (0.29)	2 (0.48)	2 (0.19)	97.56
K-10 Words :				
K-11 Words :	1 (0.29)	1 (0.24)	1 (0.09)	97.65
K-12 Words :	1 (0.29)	1 (0.24)	1 (0.09)	97.74
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	11 (2.63)	14 (1.31)	99.05
Total (unrounded)	340+?	419 (100)	1072 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1072
Different words (types):	419
Type-token ratio:	0.39

Tokens per type: 2.56

Pertaining to onlist only

Tokens:	1058
Types:	408
Families:	340
Tokens per Family :	3.11
Types per Family :	1.20

Current profile (token %)	
K-1 (82.74)	82.74
K-2 (8.02)	90.76
K-3 (3.54)	94.30
K-4 (0.84)	95.14
K-5 (1.49)	96.63
K-6 (0.28)	96.91
K-7 (0.37)	97.28
K-8 (0.09)	97.37
K-9 (0.19)	97.56
K-11 (0.09)	97.65
K-12 (0.09)	97.74
OFF (1.31)	≈100%

Types List [1]

type_[number of tokens]

BNC-COCA-1,000 types : [fams 247 : types 297 : tokens 919]

a [23] able [4] actually [1] again [1] allowed [1] also [1] although [1] american [19] americans [7] an [2] and [27] another [1] any [1] are [19] area [1] as [8] at [1] atlantic [1] australian [1] autumn [2] aware [1] be [14] became [1] because [1] becomes [1] been [1] believed [1] bernard [1] better [1] between [3] birth [1] bit [1] bokmal [1] born [1] both [2] break [1] bring [1] britain [4] british [17] britons [2] but [7] by [6] call [2] called [1] can [5] canadian [1] cannot [1] cause [1] caused [1] center [1] centre [1] certain [2] certainly [3] changes [1] chapter [1] charles [1] check [1] choose [2] chose [1] class [1] clear [1] close [1] closer [2] color [1] colour [1] common [2] congress [1] conversation [2] cornwallis [1] could [2] country [1] course [1] days [1] decided [1] deciding [1] departure [1] depend [1] differences [8] different [3] difficulty [1] discover [1] do [2] does [1] each [2] early [1] easily [1]

easy [2] either [1] ending [1] england [3] english [27] even [1] exactly [1]
fact [1] fairly [2] fall [2] few [1] fight [1] finally [1] first [1] firstly [1] flat [1]
for [6] form [1] forms [5] found [2] from [4] gas [1] general [1] gentle [1]
george [2] german [1] globally [1] good [1] gotta [1] greek [2] had [3]
half [1] hand [1] happen [1] have [2] having [1] hear [2] heard [1]
hearing [2] hebrew [2] history [1] holiday [1] how [2] i [3] idea [1] if [7]
imagine [2] in [27] instead [1] into [1] irish [1] is [14] it [13] keep [2]
ladies [1] language [6] later [1] lead [1] leaders [1] learner [1] leave [1]
led [1] less [1] lift [1] like [3] looks [2] lots [1] made [3] many [3]
marquess [1] matter [1] mean [1] meaning [1] men [1] minister [4]
misunderstandings [1] more [3] most [2] mother [2] much [2] nation [2]
nations [1] new [5] norwegian [2] norwegians [1] not [11] nothing [2]
notice [1] number [6] nynorsk [1] of [33] offered [1] often [1] on [7]
one [4] only [2] or [7] other [5] others [1] out [1] own [1] part [1]
particularly [1] partly [1] people [1] points [1] pop [1] president [5]
pretty [2] prime [4] probably [2] problems [1] pronunciation [3]
properly [1] question [2] quickly [1] rather [2] really [1] remember [1]
right [1] rubbish [1] said [3] say [4] secondly [1] see [1] shaw [1]
shop [1] should [5] side [4] silly [1] simply [1] so [5] some [4] speak [2]
speaker [2] speaking [3] started [1] states [1] still [3] store [1]
suggested [1] superman [1] sure [1] sweets [1] talking [1] teach [1]
tell [1] tend [1] than [5] that [19] the [43] their [7] them [12]
themselves [1] then [3] there [6] these [5] they [10] things [1] think [1]
this [1] those [1] throwing [1] to [25] too [2] travel [2] traveled [1]
traveling [1] travelled [1] travelling [1] two [6] under [1] united [1]
unless [1] up [3] us [1] use [1] used [2] usually [2] very [1] wanna [1]
war [1] was [12] washington [1] way [2] we [3] wear [1] wearing [1]
well [2] were [6] what [2] when [2] where [2] whether [1] which [6]
who [8] with [2] wondered [1] word [1] words [4] world [1] worn [3]
would [12] writer [1] writing [1] written [3] wrong [2] yorktown [1]
you [13] young [1] your [1]

BNC-COCA-2,000 types: [fams 42 : types 46 : tokens 59]

apartment [1] avoided [1] belong [1] biscuit [2] centuries [1] chat [1]
combine [1] correct [1] crown [1] cultural [1] defence [1] defense [1]
describe [1] discuss [1] divided [2] economically [1] equally [1]
eventually [2] example [3] exist [1] features [1] future [1] generation [1]
invited [1] labor [1] labour [1] languages [1] likely [1] messages [1]
mix [1] native [1] official [2] politically [1] pronounce [1] pronounced [2]
proposal [1] proud [1] soldier [1] speech [1] spelling [3] [3] standard [2]
theater [1] theatre [1] tongue [1] truck [1] warning [2]

BNC-COCA-3,000 types: [fams 31 : types 34 : tokens 37]

adopted [1] catalog [1] catalogue [1] colonist [1] colonists [2]
communicate [1] context [1] declared [1] differ [1] distinct [1]
elements [1] elevator [1] essays [1] flavor [1] flavour [1] formal [1]
funeral [1] hosi [1] independence [2] logical [1] loyalty [1] net [1]
pavement [1] petrol [1] portrait [1] republic [1] revolution [2] rivalry [1]
silent [1] succeeded [1] suspenders [1] terror [1] text [1] troops [1]

BNC-COCA-4,000 types: [fams 6 : types 6 : tokens 8]

impatient [1] lyrics [1] norms [1] surrenders [1] trousers [3] witty [1]

BNC-COCA-5,000 types: [fams 9 : types 10 : tokens 16]

candy [1] hack [1] pants [2] plough [1] plow [1] shorts [2] stockings [1]
underlined [1] vacation [1] vocabulary [5]

BNC-COCA-6,000 types: [fams 3 : types 3 : tokens 3]

aristocrats [1] dialects [1] garbage [1]

BNC-COCA-7,000 types: [fams 3 : types 4 : tokens 4]

cookie [1] cookies [1] inset [1] subway [1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

lorry [1]

BNC-COCA-9,000 types: [fams 1 : types 2 : tokens 2]

eavesdrop [1] eavesdropping [1]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

cheque [1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 9 : tokens 12]

bedder_2] bettuh_1] kaant_1] krent_2] sidewalk_1] terror_2] terruh_1]
underground_1] underpants_1] nynorsk_1] bokmal_1]

B Families list

a_25] able_4] actual_1] again_1] allow_1] also_1] although_1]
americans_1] and_27] another_1] any_1] area_1] as_8] at_1]
atlantic_1] australian_19] autumn_2] aware_1] be_66] because_1]
become_2] believe_1] bernard_2] better_1] between_3] birth_1] bit_1]
bokmal_1] born_1] both_2] break_1] bring_1] britain_2] british_2]
britons_27] but_7] by_6] call_3] can_6] canadian_1] cause_2]
centre_2] certain_5] change_1] chapter_1] charles_1] check_1]
choose_3] class_1] clear_1] close_3] colour_2] common_1]
congress_2] conversation_2] cornwallis_1] could_2] country_1]
course_1] day_1] decide_2] departure_2] depend_1] difference_8]
different_3] difficult_1] discover_1] do_3] each_2] early_1] easy_3]
either_1] end_1] end_of_list_1] england_2] english_4] even_1]
exact_1] fact_1] fair_2] fall_2] few_1] fight_1] final_1] find_2] first_2]
flat_1] for_6] form_6] from_4] gas_1] general_1] gentle_1] george_1]
german_3] get_1] globally_1] good_1] greek_1] half_1] hand_1]
happen_1] have_6] hear_5] hebrew_17] history_1] holiday_1] how_2]
i_3] idea_1] if_7] imagine_2] in_27] instead_1] into_1] irish_1] it_13]
keep_2] lady_1] language_1] late_1] lead_3] learn_1] leave_1] less_1]
lift_1] like_3] look_2] lot_1] make_3] man_1] many_3] marquess_6]
matter_1] mean_2] minister_1] more_3] most_2] mother_2] much_2]
nation_3] new_5] norwegian_1] norwegians_7] not_11] nothing_2]
notice_1] number_6] nynorsk_4] of_33] offer_1] often_1] on_7] one_4]
only_2] or_7] other_6] out_1] own_1] part_2] particular_1] people_1]

point_1] pop_1] president_1] pretty_2] prime_5] probably_2]
problem_1] pronunciation_4] proper_1] question_2] quick_1] rather_2]
really_1] remember_1] right_1] rubbish_1] say_7] second_1] see_1]
shaw_1] shop_1] should_5] side_4] silly_1] simple_1] so_5] some_4]
speak_7] spelling_3] start_1] states_3] still_3] store_1] suggest_1]
superman_1] sure_1] sweet_1] talk_1] teach_1] tell_1] tend_1] than_5]
that_20] the_43] then_3] there_6] they_30] thing_1] think_1] this_6]
throw_1] to_25] too_2] travel_6] two_6] under_1] understand_1]
united_1] unless_1] up_3] use_3] usual_2] very_1] want_1] war_1]
washington_1] way_2] we_4] wear_5] well_2] what_2] when_2]
where_2] whether_1] which_6] who_8] with_2] wonder_1] word_5]
world_1] would_12] write_5] wrong_2] yorktown_1] you_14] young_1]

BNC-COCA-2,000 Families: [fams 41 : types 45 : tokens 56]

apartment_1] avoid_1] belong_1] biscuit_2] century_1] chat_1]
combine_1] correct_1] crown_1] culture_1] defence_2] describe_1]
discuss_1] divide_2] economy_1] equal_1] eventually_2] example_3]
exist_1] feature_1] future_1] generation_1] invite_1] labour_2]
language_1] likely_1] message_1] mix_1] native_1] official_2]
politics_1] pronounce_3] propose_1] proud_1] soldier_1] speech_3]
standard_2] theatre_2] tongue_1] truck_1] warn_2]

BNC-COCA-3,000 Families: [fams 31 : types 34 : tokens 37]

adopt_1] catalogue_2] colony_3] communicate_1] context_1] declare_1]
differ_1] distinct_1] element_1] elevate_1] essay_1] flavour_2]
formal_1] funeral_1] host_1] independence_2] logic_1] loyal_1] net_1]
pave_1] petrol_1] portrait_1] republic_1] revolution_2] rival_1] silent_1]
succeed_1] suspend_1] terror_1] text_1] troop_1]

BNC-COCA-4,000 Families: [fams 6 : types 6 : tokens 8]

impatient_1] lyric_1] norm_1] surrender_1] trousers_3] wit_1]

BNC-COCA-5,000 Families: [fams 9 : types 10 : tokens 16]

candy_1] hack_1] pants_2] plough_2] shorts_2] stocking_1]
underline_1] vacation_1] vocabulary_5]

BNC-COCA-6,000 Families: [fams 3 : types 3 : tokens 3]

aristocrat_1] dialect_1] garbage_1]

BNC-COCA-7,000 Families: [fams 3 : types 4 : tokens 4]

cookie_[2] inset_[1] subway_[1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

lorry_[1]

BNC-COCA-9,000 Families: [fams 1 : types 2 : tokens 2]

eavesdrop_[2]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 1]

cheque_[1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 9 : tokens 12]

bedder_[2] bettuh_[1] kaant_[1] krent_[2] sidewalk_[1] terror_[2] terruh_[1]
underground_[1] underpants_[1] bokmal [1] nynorsk [1]

7.2.1.2 *A Global Language*

7.2.1.2.1 Text only file

POINTS OF DEPARTURE

Sit in pairs and try to answer the following . You may have to guess !

1 How many people in the world do you think have English as their native language ?

2 Which country has most of them ?

3 In how many countries is English the majority language ?

Two languages in the world have more native speakers than English .

Which ones ?

5 Why is English more important as a global language than either of them ?

A Global Language

When Elizabeth came to the throne in 1558 there were about 6 million people in the world who spoke English . All of them could be found in the British Isles , and most of them in England . When Elizabeth came to the throne in 1952 there were fifty times as many native speakers of English in the world . The vast majority of them lived nowhere near England . They could be found in every corner of the globe , on every continent and on islands in all three major oceans .

More than half a century on , the spectacular rise of English shows no sign of ending . In fact , it looks as if it has only just started . For every native speaker of English - and there are some 375 million of them - there are now three non native speakers . Some of them are people like you , learning English because it is an important foreign language . Others live in countries where English is an official language even though it is not a mother tongue . India , Singapore and several African countries fall into this category . English has become a global language and the world's foremost lingua franca - a term used to describe a language used as a means of communication between people whose native languages are different .

Why English ?

So how did it happen ? How on earth did a Germanic dialect spoken by a few million souls on a wind blown island on the edge of the Atlantic come to be undisputed champion of the world ? It certainly can not be because it is an easy language . It is not as you have probably discovered . It has a huge vocabulary , a highly irregular grammar and spelling that simply defies belief !

No , the language's success story is , at least to start with , part of the wider success story of the country that " invented " it . And in telling it , Elizabeth accession to the throne is not a bad place to start .

Fact Box :

A Recipe for Modern English Take dialects spoken by Germanic invaders in the 5 and 6 centuries . notably the Angles and the Saxons . The words England and English are named after the Angles . Mix them together until you get Anglo Saxon also known as Old English . Season lightly with a pinch of Scandinavian from Viking settlers . Words like " they them husband " egg and window for example . Let the mixture stand until the year 1066 when the French speaking Normans beat the English at the Battle of Hastings and become the ruling class in the country . Keep the two languages separate for a while and then gradually add French vocabulary to the Germanic language , particularly choosing words which have to do with politics . law . art . dress and food. Let the language settle until around 1350 .

It was during her reign that England began " exporting " her inhabitants in a big way . Ireland was one destination . English and later Scottish families were encouraged to settle in Ireland , pushing aside the native Gaelic speaking population . In the same period England was making its first tentative steps in settling the New World . Virginia , named after the " virgin Queen " , became the first foothold of English outside the British Isles . The first aim of colonisation was profit . New lands meant new , exotic products which could be sold at home . It also meant new markets abroad for domestic products . Of course , England was not the only European country to start behaving as if the world was theirs for the taking . Spain , Portugal , France and Holland were all busy looking for profitable colonies . That is why you can find these languages spoken today in such far flung parts of the globe . In fact , if history had been a little different , any one of them could have beaten English in the race to become the leader of the pack . In which case you would perhaps be reading this in Portuguese or Spanish .

Spreading the word

One of the main reasons you are not , is the mass settlement of North America by English speakers . Virginia , and later New England , soon became the home of thousands of English , Scottish and Irish settlers . The native Indian population was sparse and disunited and could put up little resistance . By the time of the Declaration of Independence in 1776 , the population of the 13 colonies was close to 4 million . By the beginning of the 20 century the United States of America had a population of 70 million and had long since become the biggest English speaking country in the world . During the 20 century , the US quickly took over from Britain as the motor behind the spread of English as a global language . Today , with a population of 314 million there are four times as many native speakers of English in the USA as in any other nation , even though not all Americans have English as their mother tongue .

Further north French and British interests had been in conflict since the 16 century . French settlements could be found along the coast and, especially, in the province of Quebec. In 1763 French forces were finally defeated and the whole of present day Canada came under British rule. When Britain lost its American colonies in the American War of Independence , many of those who had been loyal to the crown fled north to Canada and French speakers were soon outnumbered . But Canada remains to this day a bilingual nation. Of a total population of 34 million, two-thirds are English speakers.

North America was not the only place where English spread through mass settlement . Captain Cook's voyages in the 1770s had revealed a " terra australis incognita " , an

unknown southern continent Australia . The discovery could not have come at a better time . Until the American War of Independence , Britain had sent her most hardened criminals to Virginia . With Virginia lost , Australia seemed the perfect solution . It was as far as possible from England , it was barren and inhospitable and it was empty - except for a few harmless Aborigines . The first convicts arrived in 1788 and for many years most Australians came to the country in chains . It was only in the " that " free " settlers came in large numbers , nearly all of them from Britain . By 1900 the population was 4 million . Over a century later this has grown to 21 million .

New Zealand was a fairly late developer when it comes to mass settlement . The native Maori were a warrior people and not nearly as harmless as the Australian Aborigines. It was only when a peace treaty with them was signed in 1840 that mass immigration started . By 1900 the natives numbered some 750,000 . Today the total population is about 4.4 million , about 14 being of Maori descent .

In the parts of the world we have mentioned so far English became the dominant language by sheer weight of numbers . North America , Australia and New Zealand were seen as being " empty " land to be filled up with British settlers . But this was not the situation in most parts of the British Empire . India , for example, had been a highly developed civilisation with huge cities for centuries - much longer than Britain , in fact. Neither here nor in most of Africa was mass settlement seen as an option .

So the fact that English is an official language in , for example , India , Malaysia and many African countries today is not due to the large number of English speaking settlers . On the contrary , in most parts of the Empire the British were usually a tiny minority . During the 19 century , as the Empire grew , British man power was stretched to the limit and there was no alternative but to rely on the native population . To meet the problem a new class was needed - doctors , teachers , lawyers and administrators , who understood the natives , being natives themselves , but who spoke English and who admired British culture .

When the British finally packed up and left these colonies , as they did in the course of the 20 century , it was usually people from this class who were left in charge . By now many had lost their admiration for the Empire , but one thing they still had was their English . After independence the old colonial language still had a role to play . Many former colonies were made up of many rival tribes and languages , often in conflict with each other . Using English was therefore often a happy compromise , while in international relations having the world's leading lingua franca as an official language was a great advantage .

Most of these former colonies chose to join the Commonwealth of Nations formerly called the British Commonwealth when they became independent . Once an important trade organisation , today it is mainly a forum for discussion as well as cultural and sporting exchange . The 54 member nations , who make up nearly 30 of the world's population , are very different from each other some are rich , developed countries and others are among the poorest in the world . What they all have in common is their use of English as an official language often alongside other native languages .

Who owns English ?

Today English is used in every corner of the globe and in every field of activity . 80 of the electronically stored information in the world is in English . 66 of the world's scientists read in it . It is the language of trade , technology , sport , diplomacy , aviation - and no doubt stamp collecting and knitting too . When Israeli and Palestinian officials negotiate , they do so in English . When Hydro holds a meeting in Oslo to discuss strategy , it is held in English . The proportion of the world's population able to speak English is about a quarter - and rising . Many believe it will soon rise to a half .

Of course , many of these people do not speak it very well . Many will have a strong accent and make lots of grammatical errors . Some have developed their own local variety of English . That brings us to the question : who decides what is correct ? Who owns English ? It used to be an easy question to answer . The English owned it . As the dominant nation it was Standard English often called Oxford English that was seen as being the most correct form . Britain has long since lost its leadership of the English speaking world to the United States , and today it is American English that dominates the media , the Internet and the class room . But there are many other variants of English that are alive and well - Indian English , Australian English , South African English to name a few . In some cases , like for example Jamaican English , the local variety is so different that it is difficult for other English speakers to understand .

It seems that the more people that learn English , the more the language will live a life of its own , happily ignoring the grammar text books and dictionaries . " Japlish " , for example, is the name of a new sort of English that you will find in adverts in Japan that mixes English words with Japanese sentence structure . Then there is " Singlish " , spoken in Singapore, and " Hinglish " , a mixture of Hindi and English that is often used in India and parts of London . " Spanglish " is the new term for the language spoken by some Hispanic Americans . And do not imagine that Norway , home of " The Julecalendar " , is any different . A few years ago there was an advertising campaign in Norway for prawns , with the slogan " Reiks are good " .

Language experts predict that the impact on English of all these millions of non native speakers could be dramatic . Difficult aspects of pronunciation , like the sound , could disappear . The same goes for grammatical concord " he speaks " , " they speak " . In short , a lot of the things that lead to red ink in your essays will perhaps no longer be considered wrong .

However before you start burning your grammar book , we should warn you that these changes are still a long way off , and that experts have been known to be wrong !

7.2.1.2.2 Text Changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

isn't (2), can't, there's, don't, wasn't (2)

3. Hyphenated words with hyphen removed:

Gaelic-speaking, far-flung, English-speaking, native-speakers, mid-1800s, stamp-collecting, non-native-speakers, present-day, French-speakers

4. Compound words separated:

Foothold, classroom, manpower, windblown, textbooks

5. Words (groups of letters) removed from the text analysis:

(19)th, (20)th, The numbers have remained in the text analysis.

6. Proper nouns:

African, Malaysia, Israeli, Palestinian, Virginia, Aborigines, Australians, Maori, India, Africa, Zealand, Cook's, Indian, America, North America, New England, Scottish, Irish, Portugal, France, Holland, Portuguese, Spanish, European, Spain, British, England, Ireland, Elizabeth, English, Singapore, Anglo, Saxon, Scandinavian, Viking, French, Normans, Hastings, Germanic, Angles, Hydro, Oslo, Oxford, African, Jamaican, Japlish, Singlish, Hinglish, Hindi, London, Spanglish, Hispanic, Atlantic, Ireland, Gaelic, Americans, American, Norway, USA, Quebec, Japan, Canada, Saxons

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Common Wealth of Nations, Captain, US, Native, United States of (America), New (Zealand), (British) Isles, South (African)

Note: Text related to illustrations have been included in the text analysis.

7.2.1.2.3 Text Analysis

Text analysis: A Global Language

Vocabprofiler classic

WEB VP OUTPUT FOR FILE: A Global Language (12.02 kb)

Words recategorized by user as 1k items (proper nouns etc): AFRICAN, MALAYSIA, ISRAELI, PALESTINIAN, VIRGINIA, AUSTRALIAS, ABORIGINES, AUSTRALIANS, MAORI, INDIA, AFRICA, ZEALAND, COOK'S, INDIAN, AMERICA, ENGLAND, SCOTTISH, IRISH, PORTUGAL, FRANCE, HOLLAND, PORTUGUESE, SPANISH, EUROPEAN, SPAIN, BRITISH, ENGLAND, IRELAND, ELIZABETH, ENGLISH, SINGAPORE, ANGLO, SAXONS, SCANDINAVIAN, VIKING, FRENCH, NORMANS, HASTINGS, GERMANIC, ANGLES, HYDRO, OSLO, OXFORD, AFRICAN, JAMAICAN, JAPLISH, SINGLISH, HINGLISH, HINDI, LONDON, SPANGLISH, HISPANIC, ATLANTIC, IRELAND, GAELIC, AMERICANS, AMERICAN, NORWAY, USA, QUEBEC, BRITAIN, JAPAN, JAPANESE, CANADA, SAXON (total 176 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	333	435	1741	83.90%
Function:	(958)	(46.17%)

Content:	(783)	(37.73%)
> Anglo-Sax	(422)	(20.34%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	60	70	88	4.24%
> Anglo-Sax:	(28)	(1.35%)
1k+2k	(88.14%)
AWL Words (academic):	34	38	48	2.31%
> Anglo-Sax:	()	(0.00%)
Off-List Words:	?	128	198	9.54%
	427+?	671	2075	100%

Words in text (tokens):	2075
Different words (types):	671
Type-token ratio:	0.32
Tokens per type:	3.09
Lex density (content words/total)	0.54

Pertaining to onlist only

Tokens:	1877
Types:	543
Families:	427
Tokens per family:	4.40
Types per family:	1.27
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
83.90	83.90
4.24	88.14
2.31	90.45
9.54	100.00

AWL [34:38:48] administrators alternative aspects category communication conflict conflict contrary cultural culture domestic dominant dominant dominates dramatic errors experts experts exporting finally finally global global global global globe globe globe ignoring immigration impact major

majority majority media minority option period predict proportion rely revealed role strategy structure technology text variants

Sublist 1

exporting major majority majority period role structure variants

Sublist 2

administrators aspects category cultural culture finally finally impact strategy text

Sublist 3

alternative dominant dominant dominates immigration minority proportion rely technology

Sublist 4

communication domestic errors option predict

Sublist 5

conflict conflict

Sublist 6

experts experts ignoring revealed

Sublist 7

contrary global global global global globe globe globe media

Sublist 8

dramatic

AWL types: [34:38:48] administrators_[1] alternative_[1] aspects_[1] category_[1] communication_[1] conflict_[2] contrary_[1] cultural_[1] culture_[1] domestic_[1] dominant_[2] dominates_[1] dramatic_[1] errors_[1] experts_[2] exporting_[1] finally_[2] global_[4] globe_[3] ignoring_[1] immigration_[1] impact_[1] major_[1] majority_[2] media_[1] minority_[1] option_[1] period_[1] predict_[1] proportion_[1] rely_[1] revealed_[1] role_[1] strategy_[1] structure_[1] technology_[1] text_[1] variants_[1]

AWL families: [34:38:48]

administer_[1] alternative_[1] aspect_[1] category_[1] communicate_[1] conflict_[2] contrary_[1] culture_[2] domestic_[1] dominate_[3] drama_[1] error_[1] expert_[2] export_[1] final_[2] globe_[7] ignorant_[1] immigrate_[1] impact_[1] major_[3] media_[1] minor_[1] option_[1] period_[1] predict_[1] proportion_[1] rely_[1] reveal_[1] role_[1] strategy_[1] structure_[1] technology_[1] text_[1] vary_[1]

AWL Fr non-cognate families: [families : tokens]

Vocabprofiler (VP) Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, propers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: A Global Language (12,314 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): african, malaysia, israeli, palestinian, virginia, australian, aborigines, australians, maori, india, africa, zealand, cook 's, indian, america, england, scottish, irish, portugal, france, holland, portuguese, spanish, european, spain, british, england, ireland, elizabeth, english, singapore, anglo, saxons, scandinavian, viking, french, normans, hastings, germanic, angles, hydro, oslo, oxford, african, jamaican, japlish, singlish, hinglish, hindi, london, spanglish, hispanic, atlantic, ireland, gaelic, americans, american, norway, usa, quebec, britain, japan,japanese, canada, saxon end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	370 (67.27)	463 (68.80)	1742 (83.91)	83.91
K-2 Words :	80 (14.55)	90 (13.37)	173 (8.33)	92.24
K-3 Words :	58 (10.55)	62 (9.21)	81 (3.90)	96.14
K-4 Words :	18 (3.27)	18 (2.67)	18 (0.87)	97.01
K-5 Words :	11 (2.00)	11 (1.63)	19 (0.92)	97.93
K-6 Words :	4 (0.73)	5 (0.74)	6 (0.29)	98.22
K-7 Words :	3 (0.55)	3 (0.45)	4 (0.19)	98.41
K-8 Words :	2 (0.36)	2 (0.30)	2 (0.10)	98.51
K-9 Words :	1 (0.18)	1 (0.15)	1 (0.05)	98.56
K-10 Words :	1 (0.18)	1 (0.15)	1 (0.05)	98.61
K-11 Words :				
K-12 Words :				

K-13 Words :				
K-14 Words :	1 (0.18)	1 (0.15)	1 (0.05)	98.66
K-15 Words :	1 (0.18)	1 (0.15)	2 (0.10)	98.76
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	9 (1.34)	14 (0.67)	99.43
Total (unrounded)	550+?	673 (100)	2076 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	2076
Different words (types):	673
Type-token ratio:	0.32
Tokens per type:	3.08

Pertaining to onlist only

Tokens:	2062
Types:	664
Families:	550

Tokens per Family : 3.75
Types per Family : 1.21

Current profile (token %)	
K-1 (83.91)	83.91
K-2 (8.33)	92.24
K-3 (3.90)	96.14
K-4 (0.87)	97.01
K-5 (0.92)	97.93
K-6 (0.29)	98.22
K-7 (0.19)	98.41
K-8 (0.10)	98.51
K-9 (0.05)	98.56
K-10 (0.05)	98.61
K-14 (0.05)	98.66
K-15 (0.10)	98.76
OFF (0.67)	≈100%

Types List

Highlighted words: in the glossary

BNC-COCA-1,000 types: [fams 288 : types 358 : tokens 1581]

a [49] able [1] about [4] add [1] **advertising** [1] **advertis** [1] after [3]
ago [1] all [7] along [1] also [2] among [1] an [11] and [54] answer [2]
any [3] are [15] around [1] arrived [1] art [1] as [27] at [4] bad [1]
be [12] beat [1] beaten [1] became [4] because [2] become [4] been [5]
before [1] began [1] beginning [1] behind [1] being [4] believe [1]
better [1] between [1] big [1] biggest [1] blown [1] book [1] books [1]
box [1] brings [1] burning [1] busy [1] but [6] by [10] called [2] came [5]
can [2] case [1] cases [1] certainly [1] changes [1] charge [1]
choosing [1] chose [1] cities [1] class [4] close [1] collecting [1] come [2]
comes [1] considered [1] cook [1] corner [2] could [9] countries [5]

country [6] course [3] day [2] decides [1] did [3] different [5] difficult [2]
discovered [1] discovery [1] do [5] doctors [1] doubt [1] dress [1]
during [3] each [2] earth [1] easy [2] **edge** [1] egg [1] either [1]
empty [2] ending [1] especially [1] even [2] every [5] except [1] fact [5]
fairly [1] fall [1] families [1] far [3] few [4] field [1] fifty [1] filled [1]
finally [2] find [2] first [4] following [1] food [1] foot [1] for [20] forces [1]
form [1] found [3] four [1] free [1] from [6] further [1] get [1] goes [1]
good [1] great [1] grew [1] grown [1] guess [1] had [11] half [2]
happen [1] happily [1] happy [1] hardened [1] has [6] have [13]
having [1] he [1] held [1] her [3] here [1] highly [2] history [1] hold [1]
holds [1] home [3] how [4] however [1] huge [2] husband [1] if [3]
imagine [1] important [3] in [68] information [1] interests [1] internet [1]
into [1] is [30] island [1] islands [1] it [32] its [4] join [1] just [1] keep [1]
known [2] land [1] lands [1] large [2] late [1] later [3] law [1] lead [1]
leader [1] leadership [1] leading [1] learn [1] learning [1] least [1] left [2]
let [2] life [1] like [4] little [2] live [2] lived [1] local [2] long [3] longer [2]
looking [1] looks [1] lost [4] lot [1] lots [1] made [1] main [1] mainly [1]
major [1] make [2] making [1] man [1] many [14] markets [1] may [1]
means [1] meant [2] meet [1] meeting [1] member [1] mentioned [1]
million [10] millions [1] more [5] most [9] mother [2] much [1] name [2]
named [2] nation [3] nations [2] near [1] nearly [3] needed [1] new [10]
no [5] north [5] not [14] notably [1] now [2] number [39] numbered [1]
numbers [3] of [82] off [1] often [5] old [2] on [9] once [1] one [4]
ones [1] only [5] or [1] other [6] others [2] outside [1] over [2] own [2]
owned [1] owns [2] pack [1] packed [1] pairs [1] part [1] particularly [1]
parts [5] people [8] perfect [1] perhaps [2] place [2] play [1] points [1]
poorest [1] possible [1] power [1] present [1] probably [1] problem [1]
pushing [1] put [1] quarter [1] queen [1] question [2] quickly [1] race [1]
read [1] reading [1] reasons [1] red [1] relations [1] rich [1] rise [2]
rising [1] room [1] rule [1] ruling [1] same [2] **scientists** [1] seemed [1]
seems [1] seen [3] sent [1] settle [2] settlement [4] settlements [1]
settlers [5] settling [1] several [1] short [1] should [1] shows [1] sign [1]
signed [1] simply [1] since [3] sit [1] situation [1] so [5] sold [1] some [7]
soon [3] sort [1] sound [1] south [1] speak [3] speaker [1] speakers [9]
speaking [5] speaks [1] spoke [2] spoken [5] sport [1] sporting [1]
stand [1] start [4] started [2] steps [1] still [3] stored [1] story [2]
strong [1] such [1] take [1] taking [1] teachers [1] telling [1] term [2]
than [4] that [22] the [145] their [6] theirs [1] them [12] themselves [1]
then [2] there [9] these [6] they [8] thing [1] things [1] think [1] thirds [1]
this [6] those [1] though [2] thousands [1] three [2] through [1] time [2]
times [2] to [40] today [7] together [1] too [1] took [1] total [2] try [1]
two [3] under [1] understand [1] understood [1] unknown [1] until [4]
up [5] us [2] use [1] used [5] using [1] usually [2] very [2] war [2]
was [29] way [2] we [2] weight [1] well [3] were [11] what [2] when [10]
where [2] which [5] while [2] who [10] whole [1] whose [1] why [3]

wider_1] will_5] wind_1] window_1] with_11] word_1] words_4]
world_17] would_1] wrong_2] year_1] years_2] you_11] your_2]

BNC-COCA-2,000 types: [fams 80 : types 88 : tokens 173]

accent_1] activity_1] admiration_1] admired_1] advantage_1] alive_1]
aside_1] battle_1] captain_1] centuries_2] century_7] chains_1]
champion_1] coast_1] common_1] correct_2] criminals_1] crown_1]
cultural_1] culture_1] describe_1] developed_3] developer_1]
disappear_1] discuss_1] discussion_1] dramatic_1] due_1] empire_4]
encouraged_1] example_5] exchange_1] foreign_1] harmless_2]
ignoring_1] irregular_1] knitting_1] language_22] languages_6]
lawyers_1] lightly_1] limit_1] mass_5] minority_1] mix_1] mixes_1]
modern_1] motor_1] native_13] natives_3] neither_1] non_2] nor_1]
nowhere_1] official_4] officials_1] option_1] organisation_1] peace_1]
period_1] politics_1] population_11] products_2] pronunciation_1]
recipe_1] rely_1] remains_1] resistance_1] role_1] season_1]
sentence_1] separate_1] souls_1] southern_1] spelling_1] spread_2]
spreading_1] stamp_1] standard_1] states_2] stretched_1] success_2]
technology_1] therefore_1] tiny_1] tongue_2] trade_2] united_2]
variants_1] warn_1]

BNC-COCA-3,000 types: [fams 59 : types 63 : tokens 83]

abroad_1] administrators_1] aim_1] alongside_1] alternative_1]
angles_2] aspects_1] behaving_1] belief_1] campaign_1] category_1]
civilisation_1] colonial_1] colonies_6] colonisation_1] communication_1]
compromise_1] conflict_2] continent_2] convicts_1] declaration_1]
defeated_1] domestic_1] dominant_2] dominates_1] electronically_1]
errors_1] essays_1] experts_2] exporting_1] fled_1] former_2]
formerly_1] global_4] gradually_1] impact_1] independence_4]
independent_1] inhabitants_1] international_1] invented_1] loyal_1]
majority_2] media_1] mixture_2] negotiate_1] oceans_1] predict_1]
profit_1] profitable_1] proportion_1] province_1] revealed_1] rival_1]
solution_1] strategy_1] structure_1] text_1] treaty_1] tribes_1]
undisputed_1] variety_2] vast_1] *vast majority = one collocation

BNC-COCA-4,000 types: [fams 18 : types 18 : tokens 18]

contrary_1] defies_1] departure_1] destination_1] dictionaries_1]
exotic_1] flung_1] foremost_1] forum_1] immigration_1] ink_1]
invaders_1] pinch_1] reign_1] sheer_1] spectacular_1] virgin_1]
warrior_1]

BNC-COCA-5,000 types: [fams 11 : types 11 : tokens 19]

aviation_1] commonwealth_2] descent_1] diplomacy_1] globe_3]
grammar_3] slogan_1] tentative_1] throne_3] vocabulary_2] voyages_1]

BNC-COCA-6,000 types: [fams 4 : types 5 : tokens 6]

barren_1] dialect_1] dialects_1] isles_2] sparse_1]

BNC-COCA-7,000 types: [fams 6 : types 6 : tokens

aborigines_2] bilingual_1] grammatical_2] hispanic_1] outnumbered_1]
scandinavian_1]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

accession_1] inhospitable_1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

prawns_1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

concord_1]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams 1 : types 1 : tokens 1]

incognito_1]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 2]

lingua_2]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 6]

disunited [1] franca [2] julecalendar [1] reiks [1] terra [1]

Families list

BNC-COCA-1,000 Families: [fams 288 : types 358 : tokens 1581]

a [60] able [1] about [4] add [1] advertise [2] after [3] ago [1] all [7]
along [1] also [2] among [1] and [54] answer [2] any [3] around [1]
arrive [1] art [1] as [27] at [4] bad [1] be [106] beat [2] because [2]
become [8] before [1] begin [2] behind [1] believe [1] better [1]
between [1] big [2] blow [1] book [2] box [1] bring [1] burn [1] busy [1]
but [6] by [10] call [2] can [2] case [2] certain [1] change [1] charge [1]
choose [2] city [1] class [4] close [1] collect [1] come [8] consider [1]
cook [1] corner [2] could [9] country [11] course [3] day [2] decide [1]
different [5] difficult [2] discover [2] do [8] doctor [1] doubt [1] dress [1]
during [3] each [2] earth [1] easy [2] edge [1] egg [1] either [1]
empty [2] end [1] especially [1] even [2] every [5] except [1] fact [5]
fair [1] fall [1] family [1] far [3] few [4] field [1] fill [1] final [2] find [5]
first [4] five [1] follow [1] food [1] foot [1] for [20] force [1] form [1]
four [1] free [1] from [6] further [1] get [1] go [1] good [1] great [1]
grow [2] guess [1] half [2] happen [1] happy [2] hard [1] have [31]
he [1] here [1] high [2] history [1] hold [3] home [3] how [4] however [1]
huge [2] husband [1] if [3] imagine [1] important [3] in [68] inform [1]
interest [1] internet [1] into [1] island [2] it [36] join [1] just [1] keep [1]
know [3] land [2] large [2] late [4] law [1] lead [4] learn [2] least [1]
left [2] let [2] life [1] like [4] little [2] live [3] local [2] long [5] look [2]
lose [4] lot [2] main [2] major [1] make [4] man [1] many [14] market [1]
may [1] mean [3] meet [2] member [1] mention [1] million [11] more [5]
most [9] mother [2] much [1] name [4] nation [5] near [4] need [1]
new [10] no [5] north [5] not [14] note [1] now [2] number [43] of [82]
off [1] often [5] old [2] on [9] once [1] one [5] only [5] or [1] other [8]
out [1] over [2] own [2] owned [3] pack [2] pair [1] part [6] particular [1]
people [8] perfect [1] perhaps [2] place [2] play [1] point [1] poor [1]

possible [1] power [1] present [1] probably [1] problem [1] push [1]
put [1] quarter [1] queen [1] question [2] quick [1] race [1] read [2]
reason [1] red [1] relate [1] rich [1] rise [3] room [1] rule [2] same [2]
science [1] see [3] seem [2] sell [1] send [1] settle [13] several [1]
she [3] short [1] should [1] show [1] sign [2] simple [1] since [3] sit [1]
situation [1] so [5] some [7] soon [3] sort [1] sound [1] south [1]
speak [26] sport [2] stand [1] start [6] step [1] still [3] store [1] story [2]
strong [1] such [1] take [3] teach [1] tell [1] term [2] than [4] that [23]
the [145] then [2] there [9] they [28] thing [2] think [1] this [12]
though [2] thousand [1] three [3] through [1] time [4] to [40] today [7]
together [1] too [1] total [2] try [1] two [3] under [1] understand [2]
until [4] up [5] use [7] usual [2] very [2] war [2] way [2] we [4]
weight [1] well [3] what [2] when [10] where [2] which [5] while [2]
who [11] whole [1] why [3] wide [1] will [5] wind [1] window [1] with [11]
word [5] world [17] would [1] wrong [2] year [3] you [13]

BNC-COCA-2,000 Families: [fams 80 : types 88 : tokens 173]

accent [1] active [1] admire [2] advantage [1] alive [1] aside [1] battle [1]
captain [1] century [9] chain [1] champion [1] coast [1] common [1]
correct [2] criminal [1] crown [1] culture [2] describe [1] develop [4]
disappear [1] discuss [2] drama [1] due [1] empire [4] encourage [1]
example [5] exchange [1] foreign [1] harm [2] ignore [1] knit [1]
language [28] lawyer [1] lightly [1] limit [1] mass [5] minor [1] mix [2]
modern [1] motor [1] native [16] neither [1] non [2] nor [1] nowhere [1]
official [5] option [1] organize [1] peace [1] period [1] politics [1]
population [11] product [2] pronounce [1] recipe [1] regular [1] rely [1]
remain [1] resist [1] role [1] season [1] sentence [1] separate [1] soul [1]
southern [1] spell [1] spread [3] stamp [1] standard [1] states [2]
stretch [1] success [2] technology [1] therefore [1] tiny [1] tongue [2]
trade [2] unite [2] vary [1] warn [1]

BNC-COCA-3,000 Families: [fams 59 : types 63 : tokens 83]

abroad [1] administrator [1] aim [1] alongside [1] alternative [1] angle [2]
aspect [1] behave [1] belief [1] campaign [1] category [1] civilise [1]
colony [8] communicate [1] compromise [1] conflict [2] continent [2]
convict [1] declare [1] defeat [1] dispute [1] domestic [1] dominant [2]
dominate [1] electronic [1] error [1] essay [1] expert [2] export [1]
flee [1] former [3] global [4] gradual [1] impact [1] independence [4]
independent [1] inhabit [1] international [1] invent [1] loyal [1] majority [2]
media [1] mixture [2] negotiate [1] ocean [1] predict [1] profit [2]
proportion [1] province [1] reveal [1] rival [1] solution [1] strategy [1]
structure [1] text [1] treaty [1] tribe [1] variety [2] vast [1]

BNC-COCA-4,000 Families: [fams 18 : types 18 : tokens 18]

contrary_[1] defy_[1] departure_[1] destination_[1] dictionary_[1] exotic_[1]
fling_[1] fore_[1] forum_[1] immigrate_[1] ink_[1] invade_[1] pinch_[1]
reign_[1] sheer_[1] spectacular_[1] virgin_[1] warrior_[1]

BNC-COCA-5,000 Families: [fams 11 : types 11 : tokens 19]

aviation_[1] commonwealth_[2] descent_[1] diplomacy_[1] globe_[3]
grammar_[3] slogan_[1] tentative_[1] throne_[3] vocabulary_[2] voyage_[1]

BNC-COCA-6,000 Families: [fams 4 : types 5 : tokens 6]

barren_[1] dialect_[2] isle_[2] sparse_[1]

BNC-COCA-7,000 Families: [fams 6 : types 6 : tokens 8]

aborigine_[2] bilingual_[1] grammatical_[2] hispanic_[1] outnumber_[1]
scandinavia_[1]

BNC-COCA-8,000 Families: [fams 2 : types 2 : tokens 2]

accession_[1] hospitable_[1]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 1]

prawn_[1]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

concord_[1]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

incognito_[1]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 2]

lingua_[2]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 6]

disunited_[1] franca_[2] julecalendar_[1] reiks_[1] terra_[1]

Glossary comments:

Foothold, manpower – divided in the analysis – originally categorized as off-list word

Aborigine – added to the proper nouns list

7.2.1.3 *Renaming English*

7.2.1.3.1 Text only file

Renaming English : does the world language need a new name ?

Stewart Riddle , University of Southern Queensland

Published : June 12 , 2013 - 8 : 24

English is rapidly becoming a lingua franca in international communication for commerce and trade , education , science , international relations and tourism .

It is the fastest growing language in the world , with more people speaking English than ever before . School children in India and China are learning English at a staggering rate as their countries emphasise the importance of English as a ticket to participating in the global economy .

For example , the rise of English in China is unprecedented , and has been likened to a mania , with school children as young as seven learning to speak English .

So why then do we continue to link this evolving internationalising language with a small island in Europe that once upon a time controlled the world ?

Perhaps it is about time we got rid of the “ English ” and start calling it something else – international , standard or common language ?

Not one , but many Englishes

It is important to understand that there is not one English language ; there are many . In fact , in Australia we do not even speak and write English . We actually use Standard Australian English , which is not the same English that you might find in the United Kingdom , the United States , India or China .

There are countless blends , pidgins , creoles and mixed English languages . At the same time that English is becoming the language of internationalisation , it is also becoming localised in different parts of the world as multiple world Englishes flourish .

A sociocultural perspective on language considers the impacts of regional dialects , national standards and conventions , slang , different pronunciations and the use of communication technologies such as mobile telephones , texting and email . Our use of English depends on the contexts , audiences and purposes we are using it for .

Spoken English differs from written English . There are different ways of using written English depending on the formality and genre of writing . Spelling , grammar and punctuation change depending on who is writing and for who is reading . English is an “ open source ” language , with hybrid forms appearing all over the globe as different peoples blend English together with other languages .

Some interesting points about English languages : there are more non native speakers of English than native speakers ; nearly four out of five English speaking interactions happen between non native speakers of English ; most research is shared in English language journals ; English is the number one language used on internet sites ; English is the language of international aviation ; and most literature is published in English or translated from English into other languages .

Serious concerns with English as an international language

The rise of English comes with several concerns , including questions of cultural hegemony and postcolonial criticisms . While it is easy to shrug off such criticisms with the argument that English is necessary for social mobility , economic prosperity and education , there remain many unanswered questions around the social and cultural impacts of English as a global language .

For example , the use of English in the internationalisation of research and higher education comes at a cost to local knowledge and languages , as academics in places such as Japan , China , Germany and other parts of the world compete with scholars from the United Kingdom and USA to publish in high ranking English language research journals .

Even in France , which is renowned for its cultural and linguistic protectiveness , English is gaining ground in its universities , with 83 % of French lecturers using English in their field of research .

There is a real tragedy in the loss of language diversity as English takes over , placing other languages at risk of extinction . This has been acknowledged and efforts are being made to preserve indigenous languages in places such as Papua New Guinea , Brazil and Australia .

However , is this enough ? Are we destroying more than language through the rise of English as the international standard ?

That said , there is some sadness in the idea that we might be the last generation of travellers who experience those amusing and sometimes awkward moments when attempting to order food or ask for directions in a country where everyone does not speak English .

Stewart Riddle does not work for, consult to, own shares in or receive funding from any company or organisation that would benefit from this article, and has no relevant affiliations.

This article was originally published at The Conversation . Read the original article .

This story was found at :

7.2.1.3.2 Text Changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

Do not (1), doesn't (1),

3. Hyphenated words with hyphen removed:

non-native, English-speaking, English-language, high-ranking,

4. Compound words separated:

5. Words (groups of letters) removed from the text analysis:

AM (time of day),

6. Proper nouns:

English, Queensland, India, China, Europe, Englishes, Australia, Australian, Japan, Germany, USA, France, French, Papua New Guinea, Brazil, Stewart, Riddle

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Southern (Queensland), Standard (Australian) English, United Kingdom, United States, The Conversation,

Note: Text related to illustrations have been included in the text analysis.

7.2.1.3.3 Text Analysis

Text Analysis: Brisbane Times – Renaming English

1. VP-Classic

WEB VP OUTPUT FOR FILE: Brisbane Times - Renaming English

(4.70 kb)

Words recategorized by user as 1k items (proper nouns etc): STEWART, RIDDLE, ENGLISH, QUEENSLAND, INDIA, CHINA, EUROPE, ENGLISHES, AUSTRALIA, AUSTRALIAN, JAPAN, GERMANY, USA, FRANCE, FRENCH, PAPUA NEW GUINEA, BRAZIL (total 67 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	177	209	594	80.27%
Function:	(320)	(43.24%)
Content:	(274)	(37.03%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(140)	(18.92%)
K2 Words (1001-2000):	24	25	32	4.32%
> Anglo-Sax:	(5)	(0.68%)
1k+2k	(84.59%)
AWL Words (academic):	29	32	43	5.81%
> Anglo-Sax:	(1)	(0.14%)
Off-List Words:	<u>2</u>	<u>60</u>	<u>71</u>	<u>9.59%</u>
	230+?	326	740	100%

Current profile	
%	Cumul.
80.27	80.27
4.32	84.59
5.81	90.40
9.59	100.0

Words in text (tokens):	740
Different words (types):	326
Type-token ratio:	0.44
Tokens per type:	2.27
Lex density (content words/total)	0.57

Pertaining to onlist only

Tokens:	669
Types:	266
Families:	230
Tokens per family:	2.91

Types per family:	1.16
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

A. AWL Tokens list

AWL [29:32:43] academics acknowledged benefit communication communication consult contexts conventions cultural cultural cultural diversity economic economy emphasise evolving funding generation global global globe impacts impacts interactions journals journals lecturers link participating perspective publish published published published regional relevant research research research research sites source unprecedented

Sublist 1

benefit contexts economic economy research research research research source

Sublist 2

cultural cultural cultural impacts impacts journals journals participating regional relevant sites

Sublist 3

conventions emphasise funding interactions link publish published published published

Sublist 4

communication communication

Sublist 5

academics consult evolving generation perspective

Sublist 6

acknowledged diversity lecturers unprecedented

Sublist 7

global global globe

B. AWL Types list

AWL types: [29:32:43] academics_[1] acknowledged_[1] benefit_[1] communication_[2] consult_[1] contexts_[1] conventions_[1] cultural_[3] diversity_[1] economic_[1] economy_[1] emphasise_[1] evolving_[1] funding_[1] generation_[1] global_[2] globe_[1] impacts_[2] interactions_[1] journals_[2] lecturers_[1] link_[1] participating_[1] perspective_[1]

publish_[1] published_[3] regional_[1] relevant_[1] research_[4] sites_[1]
source_[1] unprecedented_[1]

C. AWL Families list

academy_[1] acknowledge_[1] benefit_[1] communicate_[2] consult_[1]
context_[1] convene_[1] culture_[3] diverse_[1] economy_[2] emphasis_[1]
evolve_[1] fund_[1] generation_[1] globe_[3] impact_[2] interact_[1]
journal_[2] lecture_[1] link_[1] participate_[1] perspective_[1] precede_[1]
publish_[4] region_[1] relevant_[1] research_[4] site_[1] source_[1]

AWL Fr non-cognate families: [families 1 : tokens 1] acknowledge_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: Brisbane Times - Remaining English
(4,850 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): stewart, riddle, english, queensland, india, china, europe, englishes, australia, australian, japan, germany, usa, france, french, papua new guinea, brazil, end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	176 (61.97)	205 (62.31)	572 (76.78)	76.78
K-2 Words :	48 (16.90)	53 (16.11)	90 (12.08)	88.86
K-3 Words :	39 (13.73)	43 (13.07)	56 (7.52)	96.38
K-4 Words :	7 (2.46)	7 (2.13)	7 (0.94)	97.32
K-5 Words :	5 (1.76)	5 (1.52)	5 (0.67)	97.99
K-6 Words :	3 (1.06)	3 (0.91)	3 (0.40)	98.39
K-7 Words :	2 (0.70)	2 (0.61)	2 (0.27)	98.66
K-8 Words :	1 (0.35)	1 (0.30)	1 (0.13)	98.79
K-9 Words :	1 (0.35)	1 (0.30)	1 (0.13)	98.92
K-10 Words :				

K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	2 (0.70)	2 (0.61)	2 (0.27)	99.19
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	5 (1.52)	5 (0.67)	99.86
Total (unrounded)	284+?	329 (100)	745 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	745
Different words (types):	329
Type-token ratio:	0.44
Tokens per type:	2.26

Pertaining to onlist only

Tokens:	740
Types:	324
Families:	284
Tokens per Family :	2.61
Types per Family :	1.14

Current profile (token %)	
K-1 (76.78)	76.78
K-2 (12.08)	88.86
K-3 (7.52)	96.38
K-4 (0.94)	97.32
K-5 (0.67)	97.99
K-6 (0.40)	98.39
K-7 (0.27)	98.66
K-8 (0.13)	98.79
K-9 (0.13)	98.92
K-15 (0.27)	99.19
OFF (0.67)	≈100%

A. Types List

BNC-COCA-1,000 types: [fams 146 : types 169 : tokens 509]

a [12] about [2] actually [1] all [1] also [1] an [2] and [27] any [1] appearing [1] are [8] around [1] as [14] ask [1] at [6] be [1] becoming [3] been [2] before [1] being [1] between [1] but [1] calling [1] change [1] children [2] comes [2] company [1] concerns [2] considers [1] continue [1] controlled [1] conversation [1] cost [1] countless [1] countries [1] country [1] depending [2] depends [1] different [4] do [2] does [3] easy [1] education [3] else [1] enough [1] even [2] ever [1] everyone [1] experience [1] fact [1] fastest [1] field [1] find [1] five [1] food [1] for [9] forms [1] found [1] four [1] from [5] got [1] ground [1] growing [1] happen [1] has [3] high [1] higher [1] however [1] idea [1] important [1] in [23] interesting [1] internet [1] into [1] is [23] island [1] it [7] its [2] kingdom [2] last [1] learning [2] local [1] localised [1] made [1] many [3] might [2] moments [1] more [3] most [2] name [1] national [1] nearly [1] necessary [1] need [1] new [2] no [1] not [6] number [6] of [27] off [1] on [5] once [1] one [3] open [1] or [6] order [1] other [4] our [1] out [1] over [2] own [1] papua [1] parts [2] people [1] peoples [1] perhaps [1] places [2] placing [1] points [1] protectiveness [1] questions [2] rate [1] read [1] reading [1] real [1] relations [1] renaming [1] rid [1] rise [3] sadness [1] said [1] same [2] school [2] science [1] serious [1] seven [1] several [1] shared [1] shares [1] small [1] so [1] some [2] something [1] sometimes [1] speak [3] speakers [3] speaking [2] spoken [1] start [1] story [1] such [4] takes [1] telephones [1] than [3] that [8] the [35] their [2] then [1] there [8] this [6] those [1] through [1] time [3] to [11] together [1] travellers [1] unanswered [1] understand [1] upon [1] use [4] used [1] using [3] was [2] ways [1] we [7] when [1] where [1] which [2] while [1]

who [3] why [1] with [10] work [1] world [6] would [1] write [1] writing [2] written [2] you [1] young [1]

BNC-COCA-2,000 types: [fams 48 : types 52 : tokens 90]

amusing [1] argument [1] article [3] attempting [1] awkward [1] benefit [1] commerce [1] common [1] cultural [3] destroying [1] directions [1] economic [1] economy [1] efforts [1] email [1] example [2] funding [1] gaining [1] generation [1] including [1] june [1] knowledge [1] language [16] languages [7] loss [1] mixed [1] native [3] non [2] organisation [1] original [1] originally [1] pronunciations [1] purposes [1] rapidly [1] receive [1] regional [1] remain [1] research [4] risk [1] sites [1] social [2] southern [1] spelling [1] standard [3] standards [1] states [1] technologies [1] ticket [1] tourism [1] trade [1] united [3] universities [1] university [1]

BNC-COCA-3,000 types: [fams 39 : types 41 : tokens 56]

academics [1] acknowledged [1] audiences [1] blend [1] blends [1] communication [2] compete [1] consult [1] contexts [1] conventions [1] criticisms [2] differs [1] diversity [1] emphasise [1] evolving [1] formality [1] global [2] impacts [2] importance [1] interactions [1] international [6] internationalisation [2] journals [2] lecturers [1] link [1] literature [1] mobile [1] mobility [1] multiple [1] participating [1] perspective [1] preserve [1] prosperity [1] publish [1] published [3] ranking [1] relevant [1] scholars [1] shrug [1] source [1] tragedy [1] translated [1] unprecedented [1]

BNC-COCA-4,000 types: [fams 7 : types 7 : tokens 7]

affiliations [1] flourish [1] genre [1] hybrid [1] indigenous [1] linguistic [1] staggering [1]

BNC-COCA-5,000 types: [fams 5 : types 5 : tokens 5]

aviation [1] extinction [1] globe [1] grammar [1] renowned [1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 5]

dialects [1] hegemony [1] punctuation [1] riddle [2]

BNC-COCA-7,000 types: [fams 3 : types 3 : tokens 3]

creoles [1] guinea [1] likened [1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

slang_[1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

mania_[1]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 2 : types 2 : tokens

lingua_[1] pidgins_[1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 5]

franca_[1] internationalising_[1] postcolonial_[1] sociocultural_[1] texting_[1]

B. Family List

BNC-COCA-1,000 Families: [fams 146 : types 169 : tokens 509]

a_[14] about_[2] actual_[1] all_[1] also_[1] and_[27] answer_[1] any_[1] appear_[1] around_[1] as_[14] ask_[1] at_[6] be_[37] become_[3] before_[1] between_[1] but_[1] call_[1] change_[1] child_[2] come_[2] company_[1] concern_[2] consider_[1] continue_[1] control_[1] conversation_[1] cost_[1] count_[1] country_[2] depend_[3] different_[4] do_[5] easy_[1] educate_[3] else_[1] enough_[1] even_[2] ever_[1] every_[1] experience_[1] fact_[1] fast_[1] field_[1] find_[2] five_[1] food_[1] for_[9] form_[1] four_[1] from_[5] get_[1] ground_[1] grow_[1] happen_[1] have_[3] high_[2] however_[1] idea_[1] important_[1] in_[23] interest_[1] internet_[1] into_[1] island_[1] it_[9] king_[2] last_[1] learn_[2] local_[2] make_[1] many_[3] might_[2] moment_[1] more_[3] most_[2] name_[2] nation_[1] near_[1] necessary_[1] need_[1] new_[2] no_[1] not_[6] number_[6] of_[27] off_[1] on_[5] once_[1] one_[3] open_[1] or_[6] order_[1] other_[4] out_[1] over_[2] own_[1] part_[2] people_[2] perhaps_[1] place_[3] point_[1] protect_[1] question_[2] rate_[1] read_[2] real_[1] relate_[1] rid_[1] rise_[3] sad_[1] same_[2] say_[1] school_[2] science_[1] serious_[1] seven_[1] several_[1] share_[2] small_[1] so_[1] some_[4] speak_[9] start_[1] story_[1] such_[4] take_[1] telephone_[1] than_[3] that_[9] the_[35] then_[1] there_[8] they_[2] this_[6] through_[1] time_[3] to_[11] together_[1] travel_[1] understand_[1] upon_[1] use_[8] way_[1] we_[8] when_[1] where_[1] which_[2] while_[1] who_[3] why_[1] with_[10] work_[1] world_[6] would_[1] write_[5] you_[1] young_[1]

BNC-COCA-2,000 Families: [fams 48 : types 52 : tokens 90]

amuse_[1] argue_[1] article_[3] attempt_[1] awkward_[1] benefit_[1] commerce_[1] common_[1] culture_[3] destroy_[1] direction_[1] economy_[2] effort_[1] email_[1] example_[2] fund_[1] gain_[1] generation_[1] include_[1] june_[1] knowledge_[1] language_[23] loss_[1] mix_[1] native_[3] non_[2] organize_[1] original_[2] pronounce_[1] purpose_[1] rapid_[1] receive_[1] region_[1] remain_[1] research_[4] risk_[1] site_[1] social_[2] southern_[1] spell_[1] standard_[4] states_[1] technology_[1] ticket_[1] tour_[1] trade_[1] unite_[3] university_[2]

BNC-COCA-3,000 Families: [fams 39 : types 41 : tokens 56]

academy_[1] acknowledge_[1] audience_[1] blend_[2] communicate_[2] compete_[1] consult_[1] context_[1] convention_[1] criticism_[2] differ_[1] diverse_[1] emphasise_[1] evolve_[1] formal_[1] global_[2] impact_[2] importance_[1] interact_[1] international_[8] journal_[2] lecture_[1] link_[1] literature_[1] mobile_[2] multiple_[1] participate_[1] perspective_[1] precede_[1] preserve_[1] prosper_[1] publish_[4] rank_[1] relevant_[1] scholar_[1] shrug_[1] source_[1] tragedy_[1] translate_[1]

BNC-COCA-4,000 Families: [fams 7 : types 7 : tokens 7]

affiliate_[1] flourish_[1] genre_[1] hybrid_[1] indigenous_[1] linguistic_[1] stagger_[1]

BNC-COCA-5,000 Families: [fams 5 : types 5 : tokens 5]

aviation_[1] extinct_[1] globe_[1] grammar_[1] renown_[1]

BNC-COCA-6,000 Families: [fams 4 : types 4 : tokens 5]

dialect_[1] hegemony_[1] punctuate_[1] riddle_[2]

BNC-COCA-7,000 Families: [fams 3 : types 3 : tokens 3]

creole_[1] guinea_[1] liken_[1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

slang_[1]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 1]

mania_[1]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 2 : types 2 : tokens 2]

lingua_[1] pidgin_[1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 5]

franca_[1] internationalising_[1] postcolonial_[1] sociocultural_[1] texting_[1]

7.2.1.4 Native Americans – Original Inhabitants

7.2.1.4.1 Text only file

POINTS OF DEPARTURE

You have probably seen television documentaries where a film team visits primitive tribes living deep in the jungle , far from civilization . But what do these words " primitive " and " civilized " really mean ? If we describe one society as more " advanced " than another , what are we really saying ? Does it mean that :

they look after their old people ? most people can read and write ?

nobody goes hungry ?

they have a high level of technology ?

they take care of their environment ?

they can defend themselves ?

they can wipe out their enemies ?

they are happy ?

they produce great works of art ?

they have a good understanding of science ?

they believe in God ?

they wear clothes ?

they are healthy ?

they know what is going on in the world ?

Sit in pairs / small groups and discuss which of these features are most important for a society to qualify as " advanced " . Try to agree on the five most important ones . (You can include features that are not on the list .) According to your list , do you live in an " advanced society " ?

Native Americans -

The Original Inhabitants

Where did the first inhabitants of the Americas come from ? Today we believe that they most probably migrated over the Bering Strait and settled on the American continent sometime between 20,000 and 35,000 years ago . By 10,000 BC , much of North and South America had been settled . We do not know a great deal about these early inhabitants except that they were a diverse group of peoples with more than 500 different languages . Some were nomadic tribes who followed the animals that they hunted in order to survive . As the mammoth died out , the bison or buffalo took its place and became the main source of food and hides for North American tribes . Gradually , however , some tribes made attempts

to grow food on the land , and by 300 BC traces of early village life appeared in the river valleys of New Mexico and Arizona .

Anasazi cliff dwellings . Mesa Verde . Colorado

In other words , pre Columbian America was far from being an empty wilderness free for the taking . Before Columbus arrived in the " new world " various Native American civilizations throughout North America had already developed , thrived and then mysteriously disappeared . One of these civilizations, the Anasazi , or Puebloan Native Americans as they are now referred to , built stone and adobe pueblos (villages) around the year 900 AD . Remains of these dwellings can still be seen today in the

cliff palace of Mesa Verde , Colorado . This is an amazing apartment like structure which originally had over 200 rooms built along a cliff face .

When Columbus " discovered " America in 1492 he mistakenly thought he had reached India and called these original inhabitants " Indian " . Today the preferred name is Native Americans .

Contact with Europeans

European settlement of the new world had disastrous results on the native populations . Almost from the first contact , Native American way of life was threatened . The new settlers brought new diseases against which many Native Americans had no immunity . Smallpox , in particular , killed off whole Native American communities . In fact , the drastic decline of the Native American population in the 1600s was due more to disease than to wars or armed conflicts with the new settlers .

In addition to diseases , European settlers also brought with them guns , alcohol , horses and different religious beliefs . All of these contributed to fundamentally changing Native Americans' way of life . Guns and horses changed their way of hunting for food . Attempts to convert Native Americans to Christianity undermined their spiritual beliefs . And the white man's belief that land could be owned and that others could be banned from it came into conflict with Native American beliefs that land belonged to everyone and no one in particular .

Early contact between the new settlers and the native population was sometimes friendly , but often times hostile . Although Native Americans benefited from access to new technology and trade , their very existence was seriously threatened by the new comers' thirst for land . By 1640 European settlements were already well established along the New England coast and the original inhabitants were forced to move ever westwards .

All in all , the growth of the new American nation was at the expense of the existing tribal nations . Armed conflicts with the settlers usually resulted in Native American defeat and loss of land . As more and more settlers moved into the back woods regions of the eastern colonies , Native American life was disrupted . Hunting became more difficult , forcing tribes to either go hungry , go to war or move out . As eastern tribes moved west they came into conflict with western tribes who were already there .

At the same time that Native Americans were steadily losing their land in the east , colonization of the southwestern part of the United States

was also taking place . By 1540 the Spanish had taken control of over 100 Native American pueblos in the area that is today Arizona and New Mexico , using both the sword and the cross . They forced the Native Americans there to work as slaves on their own lands and tried to convert them to Christianity . In 1680 , Pueblo Native Americans successfully rose up against the Spanish missionaries , killing Spanish priests and over 400 Spaniards . This Pueblo Revolt was a short lived victory as the Spanish regained control a dozen years later and the Pueblo Native Americans once again came under Spanish rule .

Loss of land and relocation

Defeat and loss of land , either by force or trickery , continued . As the American nation expanded , Native Americans were often forced to relocate . In 1838 - 1839 , Cherokee men women and children were removed from their homes in the area of Georgia , Tennessee and North Carolina and relocated in Oklahoma . This march has since come to be known as The Trail of Tears as it has been estimated that over 4000 men women and children , or nearly one fifth of the Cherokee nation , died during this cruel and inhumane march . Unfortunately this was not the only example of displacement .

Battles between the original inhabitants and the new settlers were frequent . Only rarely did the Native American tribes win . In one notable exception , however , at the Battle of The Little Big Horn in Southern

Montana in 1876 , Native Americans won a crushing victory . Lieutenant Colonel George Armstrong Custer , a glory hunting military leader , disobeyed orders and took his army of 650 soldiers on a foolhardy attack against the forces of six Native American tribes . It has been estimated that there were anywhere between ten and fifteen thousand Native Americans with over 2,500 warriors present in the largest concentration of Native American tribes that history has ever recorded . Not one single soldier in Custer's cavalry survived . This was , however , a short-lived victory for the Native Americans as the defeat of Custer caused public opinion to turn radically against them and it became a priority to defeat the " redskins " at any cost .

The final defeat of the Native Americans occurred in 1890 at the massacre of Wounded Knee in South Dakota when soldiers entered the Native American camp at Wounded Knee . One gun went off and uncontrolled shooting began . In the panic to get away , soldiers opened fire on men , women and children . In less than an hour , 150 Native Americans had been killed and 50 more wounded . In comparison , army casualties were 25 killed and 39 wounded . This is considered the last battle between white soldiers and Native Americans .

Robert Ottokor Lindneux (1871 - 1970) : " The Trail of Tears

Native Americans today

Although Native Americans were the original inhabitants of the American continent , they did not become American citizens until 1924 and it took another 20 years before they received the right to vote . Today there are about 330 reservations in the United States and approximately 565 federally recognized tribes or nations . Many Native Americans still live on reservations but there are many who do not . According to the 2010 census , Native Americans make up 1.7% of the entire population in the United States . Although inequities still exist today in American society , Native Americans are proud of their culture and traditions and make great efforts to preserve their heritage .

7.2.1.4.2 Text Changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

* features has been translated in the text as such: features (trekk) - trekk is taken out of the text analysis.

2. **Contractions that are written out:**

newcomers, backwoods

3. **Hyphenated words with hyphen removed:**

Pre- Columbian, apartment-like, short-lived, glory-hunting,

4. **Compound words separated:**

5. **Words (groups of letters) removed from the text analysis:**

6. **Proper nouns:**

Americas, Bering, American, Americans, America, New Mexico, Arizona, Anazazi, Mesa, Verde, Colorado, Columbus, Puebloan, India, Indian, Europeans, European, Smallpox, Christianity, England, Spanish, Pueblo, Georgia , Tennessee, Carolina , Oklahoma, Cherokee, Montana, George, Armstrong, Custer, Dakota, Robert, Ottokor, Lindneux, Spaniards,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Native Americans, (Bering) Strait, North and South America, New (Mexico), New (England), United States, (Pueblo) Revolt, North (Carolina), The Trail of Tears, Battle of The Little Big Horn, Southern

(Montana), Lieutenant Colonel (George Armstrong Custer), Wounded Knee, South (Dakota)

Note: Text related to illustrations have been included in the text analysis.

7.2.1.4.3 Text Analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Access - Native Americans Original (8.37 kb)

Words recategorized by user as 1k items (proper nouns etc): AMERICAS, BERING, AMERICAN, AMERICANS, AMERICA, NEW MEXICO, ARIZONA, ANAZAZI, MESA, VERDE, COLORADO, COLUMBUS, PUEBLOAN, INDIA, INDIAN, EUROPEANS, EUROPEAN, SMALLPOX, CHRISTIANITY, ENGLAND, SPANISH, PUEBLO, GEORGIA , TENNESSEE, CAROLINA , OKLAHOMA, CHEROKEE, MONTANA, GEORGE, ARMSTRONG, CUSTER, DAKOTA, ROBERT, OTTOKOR, LINDNEUX, SPANIARDS (total 110 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	281	360	1077	78.90%
Function:	(555)	(40.66%)
Content:	(522)	(38.24%)
> Anglo-Sax	(258)	(18.90%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	44	48	73	5.35%
> Anglo-Sax:	(29)	(2.12%)
1k+2k		(84.25%)
AWL Words (academic):	36	40	49	3.59%
> Anglo-Sax:	(4)	(0.29%)
Off-List Words:	?	92	166	12.16%
	361+?	540	1365	100%

Words in text (tokens):	1365
Different words (types):	540
Type-token ratio:	0.40
Tokens per type:	2.53
Lex density (content words/total)	0.59

Pertaining to onlist only

Tokens:	1199
Types:	448
Families:	361
Tokens per family:	3.32
Types per family:	1.24
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
78.90	78.90
5.35	84.25
3.59	87.84
12.16	100.00

A. AWL Tokens lists

AWL [36:40:49] access approximately area area benefited communities concentration conflict conflict conflicts conflicts contact contact contact contributed convert convert culture decline displacement diverse environment established estimated estimated expanded features features final fundamentally migrated military occurred priority radically regions relocate relocated relocation removed source structure survive survived team technology technology traces traditions

Sublist 1

area area benefited environment established estimated estimated occurred source structure

Sublist 2

communities culture features features final regions traditions

Sublist 3

contributed relocate relocated relocation removed technology technology

Sublist 4

access approximately concentration

Sublist 5

conflict conflict conflicts conflicts contact contact contact decline expanded fundamentally

Sublist 6

diverse migrated traces

Sublist 7

convert convert priority survive survived

Sublist 8

displacement radically

Sublist 9

military team

B. AWL Types list

AWL types: [36:40:49] access_[1] approximately_[1] area_[2] benefited_[1] communities_[1] concentration_[1] conflict_[2] conflicts_[2] contact_[3] contributed_[1] convert_[2] culture_[1] decline_[1] displacement_[1] diverse_[1] environment_[1] established_[1] estimated_[2] expanded_[1] features_[2] final_[1] fundamentally_[1] migrated_[1] military_[1] occurred_[1] priority_[1] radically_[1] regions_[1] relocate_[1] relocated_[1]

relocation_[1] removed_[1] source_[1] structure_[1] survive_[1] survived_[1]
 team_[1] technology_[2] traces_[1] traditions_[1]

C. AWL Families list

AWL families: [36:40:49]

access_[1] approximate_[1] area_[2] benefit_[1] community_[1]
 concentrate_[1] conflict_[4] contact_[3] contribute_[1] convert_[2]
 culture_[1] decline_[1] displace_[1] diverse_[1] environment_[1]
 establish_[1] estimate_[2] expand_[1] feature_[2] final_[1] fundamental_[1]
 locate_[3] migrate_[1] military_[1] occur_[1] priority_[1] radical_[1]
 region_[1] remove_[1] source_[1] structure_[1] survive_[2] team_[1]
 technology_[2] trace_[1] tradition_[1]

AWL Fr non-cognate families: [families 3 : tokens 4] feature_[2] remove_[1]
 team_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barghant)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	287 (63.78)	352 (65.06)	1059 (77.58)	77.58
K-2 Words :	90 (20.00)	102 (18.85)	176 (12.89)	90.47
K-3 Words :	40 (8.89)	46 (8.50)	79 (5.79)	96.26
K-4 Words :	16 (3.56)	16 (2.96)	18 (1.32)	97.58
K-5 Words :	4 (0.89)	4 (0.74)	4 (0.29)	97.87
K-6 Words :	4 (0.89)	4 (0.74)	4 (0.29)	98.16
K-7 Words :	3 (0.67)	3 (0.55)	3 (0.22)	98.38
K-8 Words :	1 (0.22)	1 (0.18)	1 (0.07)	98.45
K-9 Words :				
K-10 Words :	3 (0.67)	4 (0.74)	7 (0.51)	98.96
K-11 Words :	1 (0.22)	1 (0.18)	1 (0.07)	99.03
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :				

K-16 Words :				
K-17 Words :	1 (0.22)	1 (0.18)	1 (0.07)	99.10
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	4 (0.74)	4 (0.29)	99.39
Total (unrounded)	450+?	541 (100)	1365 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1365
Different words (types):	541
Type-token ratio:	0.40
Tokens per type:	2.52

Pertaining to onlist only

Tokens:	1361
Types:	537
Families:	450
Tokens per Family :	3.02
Types per Family :	1.19

Current profile (token %)	
K-1 (77.58)	77.58
K-2 (12.89)	90.47
K-3 (5.79)	96.26
K-4 (1.32)	97.58

K-5 (0.29)	97.87
K-6 (0.29)	98.16
K-7 (0.22)	98.38
K-8 (0.07)	98.45
K-10 (0.51)	98.96
K-11 (0.07)	99.03
K-17 (0.07)	99.10
OFF (0.29)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 234 : types 278 : tokens 975]

a [14] about [2] ad [1] addition [1] after [1] again [1] against [4] ago [1] agree [1] all [3] almost [1] along [2] already [3] also [2] although [3] amazing [1] an [4] and [48] animals [1] another [2] any [1] anywhere [1] appeared [1] are [9] area [2] around [1] arrived [1] art [1] as [12] at [6] away [1] back [1] be [4] became [3] become [1] been [4] before [2] began [1] being [1] believe [2] between [5] big [1] both [1] brought [2] built [2] but [3] by [6] called [1] came [3] camp [1] can [5] care [1] carolina [1] caused [1] changed [1] changing [1] children [3] clothes [1] come [2] comers [1] considered [1] continued [1] control [2] cost [1] could [2] cross [1] deal [1] deep [1] did [3] died [2] different [2] difficult [1] discovered [1] do [4] does [1] during [1] early [3] east [1] either [2] empty [1] entered [1] ever [2] everyone [1] except [1] face [1] fact [1] far [2] fifteen [1] fifth [1] film [1] final [1] fire [1] first [2] five [1] followed [1] food [3] for [6] force [1] forced [3] forces [1] forcing [1] free [1] friendly [1] from [7] georgia [1] get [1] go [2] god [1] goes [1] going [1] good [1] great [3] group [1] groups [1] grow [1] growth [1] gun [1] guns [2] had [8] happy [1] has [4] have [3] he [2] healthy [1] hides [1] high [1] his [1] history [1] homes [1] horses [2] hour [1] however [3] hungry [2] hunted [1] hunting [3] if [1] important [2] in [37] into [3] is [5] it [6] its [1] killed [3] killing [1] know [2] known [1] land [8] lands [1] largest [1] last [1] later [1] leader [1] less [1] level [1] life [4] like [1] list [2] little [1] live [2] lived [2] living [1] look [1] losing [1] made [1] main [1] make [2] man [1] many [3] mean [2] men [3] mistakenly [1] more [7] most [4] move [2] moved [2] much [1] name [1] nation [3] nations [2] nearly [1] new [13] no [2] nobody [1] north [4] not [6] notable [1] now [1] number [37] numbers [1] of [45] off [2] often [2] old [1] on [10] once [1] one [7] ones [1] only [2] opened [1] or [7] order [1] orders [1] other [1] others [1] out [3] over [6] own [1] owned [1] pairs [1] part [1] particular [2] people [2] peoples [1] place [2]

points [1] present [1] probably [2] public [1] reached [1] read [1] really [2] recorded [1] right [1] river [1] rooms [1] rose [1] rule [1] same [1] saying [1] science [1] seen [2] seriously [1] settled [2] settlement [1] settlements [1] settlers [7] shooting [1] short [2] since [1] single [1] sit [1] six [1] small [1] some [2] sometime [1] sometimes [1] south [2] spaniards [1] still [3] stone [1] take [1] taken [1] taking [2] team [1] tears [2] television [1] ten [1] than [4] that [13] the [83] their [11] them [3] themselves [1] then [1] there [5] these [7] they [20] thirst [1] this [7] thought [1] thousand [1] throughout [1] time [1] times [1] to [28] today [7] took [3] tried [1] try [1] turn [1] uncontrolled [1] under [1] understanding [1] unfortunately [1] until [1] up [2] using [1] usually [1] very [1] visits [1] war [1] wars [1] was [11] way [3] we [4] wear [1] well [1] went [1] were [12] west [1] westwards [1] what [3] when [2] where [2] which [3] white [2] who [3] whole [1] win [1] with [8] women [3] won [1] woods [1] words [2] work [1] works [1] world [3] write [1] year [1] years [3] you [3] your [1]

BNC-COCA-2,000 types: [fams 90 : types 97 : tokens 176]

access [1] according [2] advanced [3] alcohol [1] apartment [1] army [2] attack [1] attempts [2] battle [2] battles [1] belonged [1] benefited [1] citizens [1] cliff [3] coast [1] communities [1] comparison [1] concentration [1] contact [3] contributed [1] cruel [1] culture [1] describe [1] developed [1] disappeared [1] discuss [1] disease [1] diseases [2] dozen [1] due [1] efforts [1] enemies [1] entire [1] environment [1] established [1] example [1] exist [1] existence [1] existing [1] expense [1] features [2] glory [1] include [1] knee [2] languages [1] loss [3] march [2] military [1] missionaries [1] mysteriously [1] native [37] occurred [1] opinion [1] original [5] originally [1] panic [1] population [3] populations [1] preferred [1] produce [1] proud [1] qualify [1] rarely [1] received [1] recognized [1] referred [1] regions [1] relocate [1] relocated [1] relocation [1] remains [1] removed [1] reservations [2] resulted [1] results [1] slaves [1] society [4] soldier [1] soldiers [4] southern [1] spiritual [1] states [3] steadily [1] successfully [1] survive [1] survived [1] sword [1] technology [2] threatened [2] traces [1] trade [1] traditions [1] trickery [1] united [3] valleys [1] various [1] village [1] villages [1] vote [1] western [1] wipe [1] wounded [4]

BNC-COCA-3,000 types: [fams 40 : types 44 : tokens 79]

approximately [1] armed [2] banned [1] belief [1] beliefs [3] civilization [1] civilizations [2] civilized [1] colonies [1] colonization [1] conflict [2] conflicts [2] continent [2] convert [2] crushing [1] decline [1] defeat [5] defend [1] disastrous [1] disrupted [1] diverse [1] eastern [2]

estimated [2] exception [1] expanded [1] federally [1] frequent [1]
fundamentally [1] gradually [1] heritage [1] hostile [1] immunity [1]
inhabitants [7] migrated [1] palace [1] preserve [1] priests [1] priority [1]
radically [1] religious [1] source [1] structure [1] trail [2] tribal [1]
tribes [11] victory [3]

BNC-COCA-4,000 types: [fams 16 : types 16 : tokens 18]

casualties [1] census [1] colonel [1] departure [1] displacement [1]
documentaries [1] dwellings [2] horn [1] inequities [1] jungle [1]
lieutenant [1] primitive [2] regained [1] thrived [1] undermined [1]
warriors [1]

BNC-COCA-5,000 types: [fams 4 : types 4 : tokens 4]

drastic [1] massacre [1] revolt [1] wilderness [1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

cavalry [1] inhumane [1] ire [1] strait [1]

BNC-COCA-7,000 types: [fams 3 : types 3 : tokens 3]

buffalo [1] mammoth [1] nomadic [1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

adobe [1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

smallpox [1]

BNC-COCA-10,000 types: [fams 3 : types 3 : tokens 7]

bison [1] disobeyed [1] pueblo [3] pueblos [2]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

foolhardy [1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens 1]

anasazi [1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 4]

bc [2] redskins [1] southwestern [1]

B. Families list

BNC-COCA-1,000 Families: [fams 234 : types 278 : tokens 975]

a [18] about [2] add [1] advertise [1] after [1] again [1] against [4]
ago [1] agree [1] all [3] almost [1] along [2] already [3] also [2]
although [3] amaze [1] and [48] animal [1] another [2] any [2] appear [1]
area [2] around [1] arrive [1] art [1] as [12] at [6] away [1] back [1]
be [46] become [4] before [2] begin [1] believe [2] between [5] big [1]
both [1] bring [2] build [2] but [3] by [6] call [1] camp [1] can [5] care [1]
cause [1] change [2] child [3] clothes [1] come [6] consider [1]
continue [1] control [3] cost [1] could [2] cross [1] deal [1] deep [1]
die [2] different [2] difficult [1] discover [1] do [8] during [1] early [3]
east [1] either [2] empty [1] end_of_list [1] enter [1] ever [2] every [1]

except_[1] face_[1] fact_[1] far_[2] film_[1] final_[1] fire_[1] first_[2] five_[3]
follow_[1] food_[3] for_[6] force_[6] fortunate_[1] free_[1] friend_[1] from_[7]
get_[1] go_[5] god_[1] good_[1] great_[3] group_[2] grow_[2] gun_[3]
happy_[1] have_[15] he_[3] health_[1] hide_[1] high_[1] history_[1]
home_[1] horse_[2] hour_[1] however_[3] hunger_[2] hunt_[4] if_[1]
important_[2] in_[37] into_[3] it_[7] kill_[4] know_[3] land_[9] large_[1]
last_[1] late_[1] lead_[1] less_[1] level_[1] life_[4] like_[1] list_[2] little_[1]
live_[5] look_[1] lose_[1] main_[1] make_[3] man_[4] many_[3] mean_[2]
mistake_[1] more_[7] most_[4] move_[4] much_[1] name_[1] nation_[5]
near_[1] new_[13] no_[2] nobody_[1] north_[4] not_[6] note_[1] now_[1]
number_[38] of_[45] off_[2] often_[2] old_[1] on_[10] once_[1] one_[8]
only_[2] open_[1] or_[7] order_[2] other_[2] out_[3] over_[6] own_[1]
owned_[1] pair_[1] part_[1] particular_[2] people_[3] place_[2] point_[1]
present_[1] probably_[2] public_[1] reach_[1] read_[1] really_[2] record_[1]
right_[1] rise_[1] river_[1] room_[1] rule_[1] same_[1] say_[1] science_[1]
see_[2] serious_[1] settle_[11] shoot_[1] short_[2] since_[1] single_[1]
sit_[1] six_[1] small_[1] some_[4] south_[2] still_[3] stone_[1] take_[7]
team_[1] tear_[2] television_[1] ten_[1] than_[4] that_[13] the_[83] then_[1]
there_[5] they_[35] think_[1] thirst_[1] this_[14] thousand_[1] through_[1]
time_[2] to_[28] today_[7] try_[2] turn_[1] under_[1] understand_[1] until_[1]
up_[2] use_[1] usual_[1] very_[1] visit_[1] war_[2] way_[3] we_[4] wear_[1]
well_[1] west_[2] what_[3] when_[2] where_[2] which_[3] white_[2] who_[3]
whole_[1] win_[2] with_[8] woman_[3] wood_[1] word_[2] work_[2] world_[3]
write_[1] year_[4] you_[4]

BNC-COCA-2,000 Families: [fams 90 : types 97 : tokens 176]

access_[1] according_[2] advance_[3] alcohol_[1] apartment_[1] army_[2]
attack_[1] attempt_[2] battle_[3] belong_[1] benefit_[1] citizen_[1] cliff_[3]
coast_[1] community_[1] compare_[1] concentrate_[1] contact_[3]
contribute_[1] cruel_[1] culture_[1] describe_[1] develop_[1] disappear_[1]
discuss_[1] disease_[3] dozen_[1] due_[1] effort_[1] enemy_[1] entire_[1]
environment_[1] establish_[1] example_[1] exist_[3] expense_[1] feature_[2]
glory_[1] include_[1] knee_[2] language_[1] locate_[3] loss_[3] march_[2]
military_[1] mission_[1] mystery_[1] native_[37] occur_[1] opinion_[1]
original_[6] panic_[1] population_[4] prefer_[1] produce_[1] proud_[1]
qualify_[1] rare_[1] receive_[1] recognize_[1] refer_[1] region_[1] remain_[1]
remove_[1] reserve_[2] result_[2] slave_[1] society_[4] soldier_[5]
southern_[1] spirit_[1] states_[3] steady_[1] success_[1] survive_[2]
sword_[1] technology_[2] threat_[2] trace_[1] trade_[1] tradition_[1] trick_[1]
unite_[3] valley_[1] various_[1] village_[2] vote_[1] western_[1] wipe_[1]
wound_[4]

BNC-COCA-3,000 Families: [fams 40 : types 44 : tokens 79]

approximate_[1] armed_[2] ban_[1] belief_[4] civilise_[4] colony_[2]
conflict_[4] continent_[2] convert_[2] crush_[1] decline_[1] defeat_[5]
defend_[1] disaster_[1] disrupt_[1] diverse_[1] eastern_[2] estimate_[2]
exception_[1] expand_[1] federal_[1] frequent_[1] fundamental_[1]
gradual_[1] heritage_[1] hostile_[1] immune_[1] inhabit_[7] migrate_[1]
palace_[1] preserve_[1] priest_[1] priority_[1] radical_[1] religious_[1]
source_[1] structure_[1] trail_[2] tribe_[12] victory_[3]

BNC-COCA-4,000 Families: [fams 16 : types 16 : tokens 18]

casualty_[1] census_[1] colonel_[1] departure_[1] displace_[1]
documentary_[1] dwell_[2] equity_[1] horn_[1] jungle_[1] lieutenant_[1]
primitive_[2] regain_[1] thrive_[1] undermine_[1] warrior_[1]

BNC-COCA-5,000 Families: [fams 4 : types 4 : tokens 4]

drastic_[1] massacre_[1] revolt_[1] wilderness_[1]

BNC-COCA-6,000 Families: [fams 4 : types 4 : tokens 4]

cavalry_[1] humane_[1] pre_[1] strait_[1]

BNC-COCA-7,000 Families: [fams 3 : types 3 : tokens 3]

buffalo_[1] mammoth_[1] nomad_[1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

adobe_[1]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 1]

smallpox_[1]

BNC-COCA-10,000 Families: [fams 3 : types 3 : tokens 7]

bison_[1] disobey_[1] pueblo_[5]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 1]

foolhardy_[1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

[anasazi_\[1\]](#)

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 4]

[bc_\[2\]](#) [redskins_\[1\]](#) [southwestern_\[1\]](#)

7.2.1.5 Aboriginal Australians

7.2.1.5.1 Text only file

Captain Cook landing with his soldiers at Botany Bay in 1770

The end of isolation

On the 28 April 1770 , in the bay of a desolate , uncharted coast line , a historic meeting took place . A group of British sailors armed with muskets came ashore in a longboat and were met by a group of black , naked men armed with spears . Polite conversation was difficult because it was the first time either side had heard each other's language . One of the sailors later wrote down what he thought the black men had shouted . " Warra warra ! " - " Go away ! "

Not a very promising opening for the meeting of two cultures, but unfortunately rather a telling one . The bay where it took place was christened Botany Bay after the rich flora . It lies close to present day Sydney , Australia . The British sailors had just arrived aboard the

Endeavour under the command of the famous explorer James Cook . The continent that they had been standing on for all of 30 seconds was claimed for their monarch , George .

The black men were members of the Iora tribe . They and their ancestors had been fishing and hunting on this shore for 30,000 years .They had seen the intruders from the shore some hours before , but had decided to ignore them by turning their backs on them . With their white skins and strange dress they were quite obviously not of this world , and ignoring them seemed to be the best way of getting rid of them . The plan had not worked . Plan B was to shake spears and shout " Warra warra ! " .

However , these intruders were not going to go away - at least , not for long . Australia was now on the map and 18 years later more ships would arrive , this time carrying convicts from the teeming back streets of London , Glasgow and Dublin . Botany Bay was to become a penal colony . The process of colonization had begun and for the Iora tribe , and all the other tribes of this far flung continent , their age of isolation was over .

The Dreamtime

It is estimated that there were some 300,000 Aboriginal Australians at this time . Not many for such a vast area about one for every 16 square kilometres , in fact . But as nomadic hunters and gatherers they covered huge areas of the continent , also those seen by the European settlers as uninhabitable . Technologically they were still in the Stone Age . Even the bow and arrow was unknown to them . So was agriculture .However , they had developed skills in tracking and stalking that put even American Indians in the shade . With their spears , their woomeras a sort of sling for throwing spears and their boomerangs they could fell anything from a lizard to a kangaroo with extraordinary precision . They also developed hunting techniques based on enormous self control . They could , for example , stand perfectly still for several hours , spear raised , waiting for an animal to emerge from a hole .

Today we tend to associate Aborigines with the " Outback " , the desert area that covers most of central Australia . But at the time of Cook's visit most Aborigines lived on the temperate coast line where present day Australia's population is centred . The first Europeans regarded them as one people , but there were actually hundreds of tribes and around 300 different languages . Life styles varied according to the landscape they inhabited . However there were certain things they had in common . One was a special relationship to the land itself. For the European settlers it was a mystery that the Aborigines had no concept of land ownership and yet were very territorial , suffering great distress when they were forced to move away from an area . They were also mystified by their apparent lack of religion - no temples , no priests , no worship of the sun and moon .

The answer to both these mysteries was that for the Aborigines the land itself was a spiritual world , linking them to their forefathers and the forces of creation . According to Aboriginal beliefs , the world started with the Dreamtime , when there were only spirits .This Dreamtime is not over , but still here , alive and accessible in the landscape . Each hill and rock ; every tree and animal , has its own power - its dreaming , as they call it - that makes it part of a spiritual as well as a physical world . To be deprived of land was much worse than being deprived of property - it meant loss of identity and spiritual death .

The trauma of colonization

With convict ships providing the British with an endless source of slave labour, there was no use for the Aborigines. They were seen as being little more than a native pest, like the dingoes and kangaroos. As for their claim to the land, it was seen as laughable. After all, where were the villages, the fields, the domesticated animals that proved their claim? Talk of forefathers and the Dreamtime had no more meaning for the British than land ownership had for the Aborigines. In the colonists' eyes these naked, nomadic blacks were a miserable race whose days were numbered. As one

wrote in: "Nothing can stay the dying away of the Aboriginal race, which Providence has allowed to hold the land until replaced by a finer race."

Official policy was that the natives should not be mistreated, but in practice they were often killed without risk of punishment. As European settlement spread they found themselves increasingly in conflict with settlers. In these conflicts the Aborigines often put up stiff resistance, using guerrilla warfare and ambushes to terrorize settlers. But there was no coordination between tribes and their weapons were no match for guns.

Historians disagree about how many Aborigines were shot and killed by whites during the period of colonization. The figure of 10,000 has been suggested as an approximate figure. But this is only a small part of the story of Aboriginal decline. Disease accounted for around 90% of the decline in the Aboriginal population. Like the Native Americans, Aborigines had little resistance to European diseases like measles, chickenpox and smallpox, and such diseases spread like wild fire. Often they would spread in advance of direct contact with Europeans, so that by the time settlers arrived the Aboriginal communities had already been destroyed.

Nowhere was the tragedy of colonization more shocking than on the island of Tasmania. When the British first established a penal colony on what was then called Van Diemen's Land, the aboriginal population is thought to have been around 4000. They had been living there for 30,000 years as nomadic hunters and fisher men. It took just 75 years of white settlement to wipe them out. They were hunted like game, poisoned like rats and rounded up and fenced in like cattle. When the last Tasmanian died in 1876 the humiliation was still not over. Her body was boiled, the flesh removed and her skeleton exhibited in a glass case in the Tasmanian Museum, where she remained until 1947.

Adapting and surviving

In main land Australia those Aborigines who survived had no alternative but to adapt to a new life style. With white settlement came sheep, cattle and rabbits, which meant that the kangaroos and other game on which the Aborigines depended moved out. The old nomadic ways became more and more difficult, even in the Outback of central Australia. Some reserves were set up in very remote areas, but many were forced to take jobs at the new cattle stations, as farm hands or servants, mostly paid in the form of food and clothing, and regarded as little more than slaves. Others moved to towns where they rapidly became an underclass, many falling prey to poverty and alcoholism.

By the beginning of the 20 century the Aboriginal population of Australia had fallen to around 30,000. When Australia was declared a self governing commonwealth in 1901 the "first Australians" had little to celebrate. As a group they were largely ignored, having no right to vote and no status as Australian citizens.

Protest and reawakening

However, the 20 century was not only a tale of defeat for Australia native population. In the 1960 partly inspired by the Civil Rights movement in the US, there was an awakening of Aboriginal activism. Students, both white and black, held so called "freedom rides" in New South Wales to show that segregation was not just an American phenomenon. Although not official policy, it was practiced locally on a big scale, with blacks excluded from white cinemas, swimming pools, pubs etc. Aboriginal organizations grew in strength and began to demand more than just civil rights. They also wanted equality in other areas - housing, health, education - where they have lagged behind white Australians.

In recent years the issue of land rights has been high on the Aboriginal agenda. The rights of an individual are one thing, and most Australians now admit that Aboriginal people have suffered discrimination. But do

they have rights as a people, so called "native title"? Other settler states, such as the USA, Canada and New Zealand, accept the idea of a native title to land. But according to British law at the time of settlement, Australia was a terra nullius land belonging to no one. In high profile court cases Aboriginal groups have contested this view - and won.

These cases have been greeted with joy by Aboriginal organizations - and with shock by Australian industrialists, especially those involved in mining and farming. Granting native rights to huge areas of land would mean "locking up the economic future of Australia", they claimed. "Is this really one Australia for all Australians?" they asked in a newspaper advertisement. The last word on the issue of native title has not yet been said and it promises to be a hotly debated issue in the future.

The "warra warra" greeting of 1770 may not have been successful. On the other hand, the white settler's dismal prophecy of a native people "dying away" has not turned out to be right either. Today's "first Australians" take pride in their roots, and Aboriginal art and languages are making a come back. Politically they are more active than ever, with some activists saying that native title is not enough - they want sovereignty.

7.2.1.5.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. **Contractions that are written out:**

3. **Hyphenated words with hyphen removed:**

present-day, far-flung, bow-and-arrow, self-control, farm-hands, self-governing, so-called, high-profile,

4. **Compound words separated:**

Coastline, lifestyles, wildfire, fishermen, mainland, backstreets, comeback

5. **Words (groups of letters) removed from the text analysis:**

28 (th) April, George (III), 20 (th) century,

6. Proper nouns:

Aboriginal, Australians, Cook, April, British, Sydney, Australia, Endeavour, James, George, Iora, London, Glasgow, Dublin, Botany, Botanx, European, American, Indians, Aborigines, Aborigine, Dreamtime, Providence, Tasmania, Van, Diemen's, Tasmanian, Outback, Wales, USA , Canada, Zealand, Americans, Europeans,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Captain (Cook), (Botany) Bay, (Botanx) Bay, Stone Age, (Van Diemen's) Land, (Tasmanian) Museum, Civil Rights movement, New South (Wales), New (Zealand),

Note: Text related to illustrations have been included in the text analysis.

7.2.1.5.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Access - Aboriginal Australians (10.16 kb)

Words recategorized by user as 1k items (proper nouns etc): ABORIGINAL, AUSTRALIANS, COOK, APRIL, BRITISH, SYDNEY, AUSTRALIA, ENDEAVOUR, JAMES, GEORGE, IORA, LONDON, GLASGOW, DUBLIN, BOTANY, BOTANX, EUROPEAN, AMERICAN, INDIANS, ABORIGINES, ABORIGINE, DREAMTIME, PROVIDENCE, TASMANIA, VAN, DIEMEN'S, TASMANIAN, OUTBACK, WALES, USA , CANADA, ZEALAND, AMERICANS, EUROPEANS (total 94 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	327	406	1338	78.52%
Function:	(782)	(45.89%)
Content:	(556)	(32.63%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(310)	(18.19%)
K2 Words (1001-2000):	69	78	97	5.69%
> Anglo-Sax:	(45)	(2.64%)
1k+2k	(84.21%)
AWL Words (academic):	51	58	69	4.05%
> Anglo-Sax:	(4)	(0.23%)
Off-List Words:	<u>2</u>	<u>119</u>	<u>200</u>	<u>11.74%</u>
	447+?	661	1704	100%

Words in text (tokens):	1704
Different words (types):	661
Type-token ratio:	0.39
Tokens per type:	2.58

Lex density (content words/total) 0.54

Pertaining to onlist only

Tokens:	1504
Types:	542
Families:	447
Tokens per family:	3.36
Types per family:	1.21
Anglo-Sax Index: <small>(A-Sax tokens + functors / onlist tokens)</small>	%
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
78.52	78.52
5.69	84.21
4.05	88.26
11.74	100.00

A. AWL Tokens lists

AWL [51:58:69] accessible adapt adapting alternative apparent approximate area area area areas areas areas areas areas civil civil co-ordination communities concept conflict conflicts contact creation cultures debated decline decline discrimination domesticated economic emerge enormous established estimated excluded exhibited granting identity ignore ignored ignoring individual involved isolation isolation issue issue issue jobs labour linking obviously period phenomenon physical policy policy precision process removed source status style styles survived surviving techniques technologically uncharted varied

Sublist 1

area area area areas areas areas areas areas concept creation economic established estimated identity individual involved issue issue issue labour period policy policy process source varied

Sublist 2

communities cultures

Sublist 3

alternative coordination excluded linking physical removed techniques technologically

Sublist 4

accessible apparent approximate civil civil debated domesticated emerge granting jobs obviously status

Sublist 5

conflict conflicts contact decline decline precision style styles

Sublist 6

discrimination ignore ignored ignoring

Sublist 7

adapt adapting isolation isolation phenomenon survived surviving

Sublist 8

exhibited uncharted

Sublist 10

enormous

B. AWL Types list

AWL types: [51:58:69] accessible_[1] adapt_[1] adapting_[1] alternative_[1] apparent_[1] approximate_[1] area_[3] areas_[4] civil_[2] co-ordination_[1] communities_[1] concept_[1] conflict_[1] conflicts_[1] contact_[1] creation_[1] cultures_[1] debated_[1] decline_[2] discrimination_[1] domesticated_[1] economic_[1] emerge_[1] enormous_[1] established_[1] estimated_[1] excluded_[1] exhibited_[1] granting_[1] identity_[1] ignore_[1] ignored_[1] ignoring_[1] individual_[1] involved_[1] isolation_[2] issue_[3] jobs_[1] labour_[1] linking_[1] obviously_[1] period_[1] phenomenon_[1] physical_[1] policy_[2] precision_[1] process_[1] removed_[1] source_[1] status_[1] style_[1] styles_[1] survived_[1] surviving_[1] techniques_[1] technologically_[1] uncharted_[1] varied_[1]

C. AWL Families list

AWL families: [51:58:69] access_[1] adapt_[2] alternative_[1] apparent_[1] approximate_[1] area_[7] chart_[1] civil_[2] community_[1] concept_[1] conflict_[2] contact_[1] coordinate_[1] create_[1] culture_[1] debate_[1] decline_[2] discriminate_[1] domestic_[1] economy_[1] emerge_[1] enormous_[1] establish_[1] estimate_[1] exclude_[1] exhibit_[1] grant_[1] identify_[1] ignorant_[3] individual_[1] involve_[1] isolate_[2] issue_[3] job_[1] labour_[1] link_[1] obvious_[1] period_[1] phenomenon_[1] physical_[1] policy_[2] precise_[1] process_[1] remove_[1] source_[1] status_[1] style_[2] survive_[2] technique_[1] technology_[1] vary_[1]

AWL Fr non-cognate families: [families 4 : tokens 4] grant_[1] involve_[1] obvious_[1] remove_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	331 (60.40)	404 (60.94)	1364 (80.00)	80.00
K-2 Words :	114 (20.80)	132 (19.91)	170 (9.97)	89.97
K-3 Words :	56 (10.22)	63 (9.50)	80 (4.69)	94.66
K-4 Words :	15 (2.74)	15 (2.26)	16 (0.94)	95.60
K-5 Words :	9 (1.64)	10 (1.51)	14 (0.82)	96.42
K-6 Words :	6 (1.09)	6 (0.90)	6 (0.35)	96.77
K-7 Words :	5 (0.91)	5 (0.75)	9 (0.53)	97.30
K-8 Words :	3 (0.55)	3 (0.45)	3 (0.18)	97.48
K-9 Words :	5 (0.91)	6 (0.90)	7 (0.41)	97.89
K-10 Words :	1 (0.18)	1 (0.15)	1 (0.06)	97.95
K-11 Words :	1 (0.18)	1 (0.15)	2 (0.12)	98.07
K-12 Words :				
K-13 Words :	1 (0.18)	1 (0.15)	1 (0.06)	98.13
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :	1 (0.18)	1 (0.15)	1 (0.06)	98.19
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	9 (1.36)	14 (0.82)	99.01

Total (unrounded) 548+? 663 (100) 1705 (100) ≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens): 1705
 Different words (types): 663
 Type-token ratio: 0.39
 Tokens per type: 2.57

Pertaining to onlist only

Tokens: 1691
 Types: 654
 Families: 548
 Tokens per Family : 3.09
 Types per Family : 1.19

Current profile (token %)	
K-1 (80.00)	80.00
K-2 (9.97)	89.97
K-3 (4.69)	94.66
K-4 (0.94)	95.60
K-5 (0.82)	96.42
K-6 (0.35)	96.77
K-7 (0.53)	97.30
K-8 (0.18)	97.48
K-9 (0.41)	97.89
K-10 (0.06)	97.95
K-11 (0.12)	98.07
K-13 (0.06)	98.13
K-17 (0.06)	98.19
OFF (0.82)	≈100%

a [36] about [2] accept [1] actually [1] admit [1] advertisement [1] after [2] age [2] all [4] allowed [1] already [1] also [4] although [1] an [8] and [52] animal [2] animals [1] answer [1] anything [1] apparent [1] are [3] area [3] areas [4] around [4] arrive [1] arrived [2] art [1] as [19] asked [1] at [6] away [5] back [2] backs [1] based [1] be [5] became [2] because [1] become [1] been [10] before [1] began [1] beginning [1] begun [1] behind [1] being [2] best [1] between [1] big [1] black [4] blacks [2] body [1] both [2] but [13] buy [1] by [11] call [1] called [3] came [2] can [1] carrying [1] case [1] cases [2] central [2] centred [1] certain [1] close [1] clothing [1] come [1] control [1] conversation [1] cook [3] could [2] court [1] covered [1] covers [1] day [2] days [1] death [1] decided [1] depended [1] died [1] different [1] difficult [2] do [1] down [1] dreaming [1] dress [1] during [1] dying [2] each [2] education [1] either [2] end [1] endless [1] enough [1] especially [1] europeans [2] even [3] ever [1] every [1] eyes [1] fact [1] fallen [1] falling [1] far [1] farm [1] farming [1] fell [1] fields [1] figure [2] finer [1] fire [1] first [5] fishing [1] food [1] for [23] forced [2] forces [1] form [1] found [1] freedom [1] from [6] game [2] getting [1] glass [1] go [2] going [1] governing [1] great [1] grew [1] group [3] groups [1] guns [1] had [20] hand [1] hands [1] has [6] have [7] having [1] he [1] health [1] heard [1] held [1] her [2] here [1] high [2] hill [1] his [1] historians [1] historic [1] hold [1] hole [1] hotly [1] hours [2] housing [1] how [1] however [4] huge [2] hundreds [1] hunted [1] hunters [2] hunting [2] idea [1] in [36] involved [1] is [7] island [1] issue [3] it [12] its [2] itself [2] jobs [1] just [4] killed [2] land [13] landing [1] largely [1] last [2] later [2] laughable [1] law [1] least [1] lies [1] life [2] like [7] line [2] little [4] lived [1] living [1] locally [1] locking [1] long [1] main [1] makes [1] making [1] many [4] may [1] mean [1] meaning [1] meant [2] meeting [2] members [1] men [4] met [1] mistreated [1] more [9] most [3] mostly [1] move [1] moved [2] movement [1] much [1] new [4] no [12] not [16] nothing [1] now [2] number [25] numbered [1] obviously [1] of [58] often [3] old [1] on [14] one [9] only [3] opening [1] or [1] other [6] others [1] out [3] over [3] own [1] ownership [2] paid [1] part [2] partly [1] people [4] perfectly [1] place [2] plan [2] power [1] present [2] promises [1] promising [1] put [2] quite [1] rabbits [1] race [3] raised [1] rather [1] really [1] recent [1] relationship [1] rich [1] rid [1] rides [1] right [2] rights [6] rock [1] rounded [1] said [1] sailors [3] saying [1] seemed [1] seen [4] self [2] servants [1] set [1] settlement [4] settler [2] settlers [5] several [1] shake [1] she [1] ships [2] shot [1] should [1] shout [1] shouted [1] show [1] side [1] skins [1] small [1] so [4] some [4] sort [1] south [1] special [1] square [1] stand [1] standing [1] started [1] stations [1] stay [1] still [4] stone [1] story [1] strange [1] streets [1] students [1] such [3] suggested [1] sun [1] swimming [1] take [2] talk [1] telling [1] tend [1] than [7] that [14] the [107] their [14] them [8] themselves [1] then [1] there [8] these [5]

A. Types list

BNC-COCA-1,000 types: [fams 278 : types 334 : tokens 1290]

they [30] thing [1] things [1] this [9] those [3] thought [2] throwing [1]
time [6] tire [1] to [39] today [2] took [3] towns [1] tracking [1] tree [1]
turned [1] turning [1] two [1] under [1] unfortunately [1] unknown [1]
until [2] up [4] us [1] use [1] using [1] van [1] very [4] view [1] visit [1]
waiting [1] want [1] wanted [1] was [29] way [1] ways [1] we [1] well [1]
were [23] what [2] when [5] where [6] which [3] white [7] whites [1]
who [1] whose [1] wild [1] with [17] without [1] won [1] word [1]
worked [1] world [4] worse [1] would [3] wrote [2] years [5] yet [2]
*wildfire, far-flung in glossary

BNC-COCA-2,000 types: [fams 115 : types 128 : tokens 171]

accessible [1] according [3] accounted [1] active [1] activism [1]
activists [1] adapt [1] adapting [1] advance [1] alcoholism [1] alive [1]
april [1] associate [1] awakening [1] bay [4] belonging [1] boiled [1]
bow [1] captain [1] century [2] citizens [1] claim [2] claimed [2] coast [2]
command [1] common [1] communities [1] contact [1] creation [1]
cultures [1] demand [1] desert [1] destroyed [1] developed [2] direct [1]
disease [1] diseases [2] economic [1] enormous [1] equality [1]
established [1] example [1] famous [1] fenced [1] future [2] gatherers [1]
granting [1] identity [1] ignore [1] ignored [1] ignoring [1] increasingly [1]
individual [1] industrialists [1] joy [1] kilometres [1] labour [1] lack [1]
language [1] languages [2] loss [1] map [1] match [1] moon [1]
mysteries [1] mystery [1] native [9] natives [1] newspaper [1]
nowhere [1] official [2] organizations [2] period [1] physical [1]
poisoned [1] policy [2] polite [1] politically [1] pools [1] population [5]
practice [1] practiced [1] pride [1] process [1] property [1] protest [1]
proved [1] providing [1] pubs [1] punishment [1] rapidly [1] rats [1]
reawakening [1] regarded [2] remained [1] removed [1] replaced [1]
reserves [1] resistance [2] risk [1] roots [1] scale [1] seconds [1]
shade [1] sheep [1] shock [1] shocking [1] shore [2] skills [1] slave [1]
slaves [1] soldiers [1] spirits [1] spiritual [3] spread [3] states [1] stiff [1]
strength [1] style [1] styles [1] successful [1] suffered [1] suffering [1]
survived [1] surviving [1] tale [1] technologically [1] title [4] varied [1]
villages [1] vote [1] weapons [1] wipe [1]

BNC-COCA-3,000 types: [fams 56 : types 62 : tokens 80]

agenda [1] agriculture [1] alternative [1] approximate [1] armed [2]
beliefs [1] cattle [3] celebrate [1] civil [2] colonist [1] colonization [4]
colony [2] concept [1] conflict [1] conflicts [1] contested [1] continent [3]
convict [1] convicts [1] coordination [1] debated [1] declared [1]
decline [2] defeat [1] disagree [1] discrimination [1] emerge [1]
estimated [1] etc [1] excluded [1] exhibited [1] explorer [1]
extraordinary [1] flesh [1] greeted [1] greeting [1] inhabited [1]

inspired [1] isolation [2] landscape [2] linking [1] mining [1] museum [1]
naked [2] phenomenon [1] poverty [1] precision [1] priests [1] profile [1]
religion [1] remote [1] source [1] sovereignty [1] status [1] techniques [1]
territorial [1] terrorize [1] tragedy [1] tribe [2] tribes [3] uncharted [1]
uninhabitable [1] vast [1]

BNC-COCA-4,000 types: [fams 16 : types 16 : tokens 17]

aboard [1] ancestors [1] arrow [1] cinemas [1] deprived [2] distress [1]
endeavour [1] flung [1] guerrilla [1] humiliation [1] miserable [1]
monarch [1] prey [1] temples [1] trauma [1] worship [1]

BNC-COCA-5,000 types: [fams 10 : types 10 : tokens 16]

botany [2] commonwealth [1] intruders [2] logged [1] pest [1]
segregation [1] skeleton [1] spear [1] spears [4] stalking [1] warfare [1]

BNC-COCA-6,000 types: [fams 6 : types 6 : tokens 6]

ambushes [1] ashore [1] dismal [1] flora [1] lizard [1] sling [1]

BNC-COCA-7,000 types: [fams 6 : types 7 : tokens 36]

aboriginal [16] aborigines [11] desolate [1] nomadic [4] penal [2]
prophecy [1] temperate [1]

BNC-COCA-8,000 types: [fams 3 : types 3 : tokens 3]

christened [1] domesticated [1] mystified [1]

BNC-COCA-9,000 types: [fams 5 : types 6 : tokens 7]

kangaroo [1] kangaroos [2] muskets [1] smallpox [1] leeming [1]
underclass [1]

BNC-COCA-10,000 types: [fams 2 : types 2 : tokens 2]

measles [1] providence [1]

BNC-COCA-11,000 types: [fams 2 : types 2 : tokens 4]

forefathers [2] outback [2]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams 1 : types 1 : tokens 1]

boomerangs_1]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens 1]

dingoes_1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams 1 : types 1 : tokens 4]

dreamtime_4]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 9 : tokens 14]

chickenpox_1] fisher_1] Inumber_1] longboat_1] nullius_1] numberl_1]
terra_1] warra_6] woomearas_1]

B. Families list

BNC-COCA-1,000 Families: [fams 278 : types 334 : tokens 1290]

a_44] about_2] accept_1] actual_1] admit_1] advertise_1] after_2]
age_2] all_4] allow_1] already_1] also_4] although_1] and_52]
animal_3] answer_1] any_1] apparent_1] area_7] around_4] arrive_3]
art_1] as_19] ask_1] at_6] away_5] back_3] base_1] be_79]
because_1] become_3] before_1] begin_3] behind_1] best_1]
between_1] big_1] black_6] body_1] both_2] but_13] buy_1] by_11]
call_4] can_1] carry_1] case_3] centre_3] certain_1] close_1]

clothes_1] come_3] control_1] conversation_1] cook_3] could_2]
court_1] cover_2] day_3] death_1] decide_1] depend_1] die_3]
different_1] difficult_2] do_1] down_1] dream_1] dress_1] during_1]
each_2] educate_1] either_2] end_2] end_of_list_2] enough_1]
especially_1] even_3] ever_1] every_1] eye_1] fact_1] fall_3] far_1]
farm_2] field_1] figure_2] find_1] fine_1] fire_1] first_5] fish_1] food_1]
for_23] force_3] form_1] fortunate_1] free_1] from_6] game_2] get_1]
glass_1] go_3] govern_1] great_1] group_4] grow_1] gun_1] hand_2]
have_34] he_2] health_1] hear_1] here_1] high_2] hill_1] history_2]
hold_2] hole_1] hot_1] hour_2] house_1] how_1] however_4] huge_2]
hundred_1] hunt_5] idea_1] in_36] involve_1] island_1] issue_3] it_16]
job_1] just_4] kill_2] know_1] land_14] large_1] last_2] late_2]
laugh_1] law_1] least_1] lie_1] life_2] like_7] line_2] little_4] live_2]
local_1] lock_1] long_1] main_1] make_2] man_4] many_4] may_1]
mean_4] meet_3] member_1] more_9] most_4] move_4] much_1]
new_4] no_12] not_16] nothing_1] now_2] number_26] obvious_1]
of_58] often_3] old_1] on_14] one_9] only_3] open_1] or_1] other_7]
out_3] over_3] own_1] owned_2] part_3] pay_1] people_4] perfect_1]
place_2] plan_2] power_1] present_2] promise_2] put_2] quite_1]
rabbit_1] race_3] raise_1] rather_1] really_1] recent_1] relate_1]
rich_1] rid_1] ride_1] right_2] rights_6] rock_1] round_1] sail_3] say_2]
see_4] seem_1] self_2] serve_1] set_1] settle_11] several_1] shake_1]
she_3] ship_2] shoot_1] should_1] shout_2] show_1] side_1] skin_1]
small_1] so_4] some_4] sort_1] south_1] special_1] square_1]
stand_2] start_1] station_1] stay_1] still_4] stone_1] store_1]
strange_1] street_1] student_1] such_3] suggest_1] sun_1] swim_1]
take_5] talk_1] tell_1] tend_1] than_7] that_17] the_107] then_1]
there_8] they_53] thing_2] think_2] this_14] throw_1] time_6] tire_1]
to_39] today_2] town_1] track_1] treat_1] tree_1] turn_2] two_1]
under_1] until_2] up_4] use_2] van_1] very_4] view_1] visit_1] wait_1]
want_2] way_2] we_2] well_1] what_2] when_5] where_6] which_3]
white_8] who_2] wild_1] win_1] with_17] without_1] word_1] work_1]
world_4] worse_1] would_3] write_2] year_5] yet_2]

BNC-COCA-2,000 Families: [fams 115 : types 128 : tokens 171]

access_1] according_3] account_1] active_3] adapt_2] advance_1]
alcohol_1] alive_1] april_1] associate_1] awake_2] bay_4] belong_1]
boil_1] bow_1] captain_1] century_2] citizen_1] claim_4] coast_2]
command_1] common_1] community_1] contact_1] create_1] culture_1]
demand_1] desert_1] destroy_1] develop_2] direct_1] disease_3]
economy_1] enormous_1] equal_1] establish_1] example_1] famous_1]
fence_1] future_2] gather_1] grant_1] identify_1] ignore_3] increase_1]
individual_1] industry_1] joy_1] kilometre_1] labour_1] lack_1]
language_3] loss_1] map_1] match_1] moon_1] mystery_2] native_10]

newspaper_[1] nowhere_[1] official_[2] organize_[2] period_[1] physical_[1]
poison_[1] policy_[2] polite_[1] politics_[1] pool_[1] population_[5]
practise_[2] pride_[1] process_[1] property_[1] protest_[1] prove_[1]
provide_[1] pub_[1] punish_[1] rapid_[1] rat_[1] regard_[2] remain_[1]
remove_[1] replace_[1] reserve_[1] resist_[2] risk_[1] root_[1] scale_[1]
seconds_[1] shade_[1] sheep_[1] shock_[2] shore_[2] skill_[1] slave_[2]
soldier_[1] spirit_[4] spread_[3] states_[1] stiff_[1] strength_[1] style_[2]
success_[1] suffer_[2] survive_[2] tale_[1] technology_[1] title_[4] vary_[1]
village_[1] vote_[1] weapon_[1] wipe_[1]

BNC-COCA-3,000 Families: [fams 56 : types 62 : tokens 80]

agenda_[1] agriculture_[1] alternative_[1] approximate_[1] armed_[2]
belief_[1] cattle_[3] celebrate_[1] chart_[1] civil_[2] colony_[7] concept_[1]
conflict_[2] contest_[1] continent_[3] convict_[2] coordinate_[1] debate_[1]
declare_[1] decline_[2] defeat_[1] disagree_[1] discriminate_[1] emerge_[1]
estimate_[1] etc_[1] exclude_[1] exhibit_[1] explore_[1] extraordinary_[1]
flesh_[1] greet_[2] inhabit_[2] inspire_[1] isolate_[2] landscape_[2] link_[1]
miner_[1] museum_[1] naked_[2] phenomenon_[1] poverty_[1] precise_[1]
priest_[1] profile_[1] religion_[1] remote_[1] source_[1] sovereign_[1]
status_[1] technique_[1] territory_[1] terror_[1] tragedy_[1] tribe_[5] vast_[1]

BNC-COCA-4,000 Families: [fams 16 : types 16 : tokens 17]

aboard_[1] ancestor_[1] arrow_[1] cinema_[1] deprive_[2] distress_[1]
endeavour_[1] fling_[1] guerilla_[1] humiliate_[1] miserable_[1] monarch_[1]
prey_[1] temple_[1] trauma_[1] worship_[1]

BNC-COCA-5,000 Families: [fams 10 : types 10 : tokens 16]

botany_[2] commonwealth_[1] intrude_[2] lag_[1] pest_[1] segregate_[1]
skeleton_[1] spear_[5] stalk_[1] warfare_[1]

BNC-COCA-6,000 Families: [fams 6 : types 6 : tokens 6]

ambush_[1] ashore_[1] dismal_[1] flora_[1] lizard_[1] sling_[1]

BNC-COCA-7,000 Families: [fams 6 : types 7 : tokens 36]

aborigine_[27] desolate_[1] nomad_[4] penal_[2] prophecy_[1]
temperate_[1]

BNC-COCA-8,000 Families: [fams 3 : types 3 : tokens 3]

christen_[1] domesticate_[1] mystify_[1]

BNC-COCA-9,000 Families: [fams 5 : types 6 : tokens 7]

kangaroo_[3] musket_[1] smallpox_[1] teem_[1] underclass_[1]

BNC-COCA-10,000 Families: [fams 2 : types 2 : tokens 2]

measles_[1] providence_[1]

BNC-COCA-11,000 Families: [fams 2 : types 2 : tokens

forefather_[2] outback_[2]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams 1 : types 1 : tokens 1]

boomerang_[1]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

dingo_[1]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams 1 : types 1 : tokens 4]

dreamtime_[4]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 9 : tokens 14]

chickenpox [1] fisher [1] lnumber [1] longboat [1] nullius [1] numberl [1]
terra [1] warra [6] woomearas [1]

7.2.1.6 Stolen Children

7.2.1.6.1 Text only file

At the beginning of the 20 century it was decided that Aborigines of mixed race should be assimilated into main stream Australian society . This led to a practice in which light skinned Aboriginal children were forcibly removed from their parents and adopted by white families . Some families even " blackened up " their children to avoid them being taken . As many as 30,000 " stolen children " , as they were called , were uprooted in this way between 1900 and 1970 .

It was only that these practices were given media attention . There was a public outcry and the issue forced white Australians to face up to the less heroic aspects of their country's past . In 1998 a National Sorry Day was held (although not supported by the federal government) to apologize for past wrongs in general and the sufferings of " the stolen generation " in particular . In 2000 , the year of the Sydney Olympics , 400,000 people took part in a " Walk of Reconciliation " across Sydney Harbour Bridge . At the games themselves the Olympic flame was lit by Aboriginal athlete Cathy Freeman .

Archie Roach is an Australian singer song writer . He was himself one of the " stolen generation " . Along with his sisters he was forcibly removed from his Aboriginal parents and placed in an orphanage . Later he was fostered by Scottish immigrants . Roach didn't learn about his origins until he was a young man . He reacted with anger and left his foster home carrying only a guitar . After living on the streets of Adelaide and Melbourne for many years he had his break through as a musician and has become a highly respected artist . He has toured in America and Europe with artists like Tracy Chapman and Bob Dylan .

This song is from his debut album Charcoal Lane (1992) .

7.2.1.6.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

light-skinned, singer-songwriter

4. Compound words separated:

Breakthrough, mainstream

5. Words (groups of letters) removed from the text analysis:

20th

6. Proper nouns:

Australian, Australians, Aborigines, Aboriginal, Sydney, Olympics, Cathy, Freeman, Archie, Roach, Scottish, Adelaide, Melbourne, America, Europe, Tracy, Chapman, Bob, Dylan

Take note: The words outside of brackets have not been placed on the list of proper nouns.

National Sorry Day, Walk of Reconciliation, (Sydney) Harbour Bridge, Charcoal Lane,

Note: Text related to illustrations have been included in the text analysis.

7.2.1.6.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Access - Stolen Children (1.76 kb)

Words recategorized by user as 1k items (proper nouns etc): AUSTRALIAN, AUSTRALIANS, ABORIGINES, ABORIGINAL, SYDNEY, OLYMPICS, CATHY, FREEMAN, ARCHIE, ROACH, SCOTTISH, ADELAIDE, MELBOURNE, AMERICA, EUROPE, TRACY, CHAPMAN, BOB, DYLAN (total 24 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	95	113	223	75.08%
Function:	(140)	(47.14%)
Content:	(83)	(27.95%)
> Anglo-Sax	(46)	(15.49%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	16	17	21	7.07%
> Anglo-Sax:	(3)	(1.01%)
1k+2k	(82.15%)
AWL Words (academic):	8	8	10	3.37%
> Anglo-Sax:	(2)	(0.67%)
Off-List Words:	<u>2</u>	<u>37</u>	<u>43</u>	<u>14.48%</u>
	119+?	175	297	100%

Words in text (tokens):	297
Different words (types):	175
Type-token ratio:	0.59
Tokens per type:	1.70
Lex density (content words/total)	0.53

Pertaining to onlist only

Tokens:	254
Types:	138
Families:	119
Tokens per family:	2.13
Types per family:	1.16
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
75.08	75.08
7.07	82.15
3.37	85.52
14.48	100.00

A. AWL Tokens lists

AWL [8:8:10] aspects federal generation generation immigrants issue media reacted removed removed

Sublist 1

issue

Sublist 2

aspects

Sublist 3

immigrants reacted removed removed

Sublist 5

generation generation

Sublist 6

federal

Sublist 7

media

B. AWL Types list

AWL types: [8:8:10] aspects_[1] federal_[1] generation_[2] immigrants_[1] issue_[1] media_[1] reacted_[1] removed_[2]

C. AWL Families list

aspect_[1] federal_[1] generation_[2] immigrate_[1] issue_[1] media_[1] react_[1] remove_[2]

AWL Fr non-cognate families: [families 1 : tokens 2] remove_[2]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	118 (76.13)	135 (76.70)	255 (85.86)	85.86
K-2 Words :	18 (11.61)	19 (10.80)	21 (7.07)	92.93
K-3 Words :	12 (7.74)	13 (7.39)	13 (4.38)	97.31
K-4 Words :	2 (1.29)	2 (1.14)	2 (0.67)	97.98
K-5 Words :	2 (1.29)	2 (1.14)	2 (0.67)	98.65
K-6 Words :	1 (0.65)	1 (0.57)	1 (0.34)	98.99
K-7 Words :				
K-8 Words :	2 (1.29)	2 (1.14)	2 (0.67)	99.66
K-9 Words :				
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				

K-25 Words :

Off-List:	??	0 (0.00)	0 (0.00)	
Total (unrounded)	155+?	176 (100)	297 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	297
Different words (types):	176
Type-token ratio:	0.59
Tokens per type:	1.69

Pertaining to onlist only

Tokens:	297
Types:	176
Families:	155
Tokens per Family :	1.92
Types per Family :	1.14

Current profile (token %)	
K-1 (85.86)	85.86
K-2 (7.07)	92.93
K-3 (4.38)	97.31
K-4 (0.67)	97.98
K-5 (0.67)	98.65
K-6 (0.34)	98.99
K-8 (0.67)	99.66
OFF (0.00)	≈100%

a [8] about [1] across [1] after [1] along [1] although [1] an [2] and [10] artist [1] artists [1] as [4] at [2] be [1] become [1] beginning [1] being [1] between [1] blackened [1] break [1] by [4] called [1] carrying [1] children [4] country [1] day [1] decided [1] did [1] dylan [1] even [1] face [1] families [2] for [2] forced [1] forcibly [2] from [3] games [1] general [1] given [1] government [1] had [1] has [2] he [7] held [1] highly [1] himself [1] his [6] home [1] in [9] into [1] is [2] issue [1] it [2] later [1] learn [1] led [1] left [1] less [1] light [1] like [1] lit [1] living [1] main [1] man [1] many [2] musician [1] national [1] not [2] number [10] of [8] on [1] one [1] only [2] parents [2] part [1] particular [1] past [2] people [1] placed [1] public [1] race [1] should [1] singer [1] sisters [1] skinned [1] some [1] song [2] sorry [1] stolen [4] streets [1] supported [1] taken [1] that [2] the [13] their [3] them [1] themselves [1] there [1] these [1] they [1] this [3] through [1] to [5] took [1] until [1] up [2] walk [1] was [9] way [1] were [4] which [1] white [2] with [3] writer [1] wrongs [1] year [1] years [1] young [1] *mainstream in glossary

BNC-COCA-2,000 types: [fams 18 : types 19 : tokens 21]

anger [1] attention [1] avoid [1] bridge [1] century [1] flame [1] generation [2] heroic [1] lane [1] mixed [1] practice [1] practices [1] reacted [1] removed [2] respected [1] society [1] stream [1] sufferings [1] toured [1]

BNC-COCA-3,000 types: [fams 12 : types 12 : tokens 13]

adopted [1] album [1] apologize [1] aspects [1] athlete [1] federal [1] foster [1] fostered [1] guitar [1] harbour [1] immigrants [1] media [1] origins [1]

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

debut [1] reconciliation [1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

assimilated [1] orphanage [1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

charcoal [1]

BNC-COCA-7,000 types: [fams 1 : types 2 : tokens 4]

aboriginal [3] aborigines [1]

A. Types list

BNC-COCA-1,000 types: [fams 90 : types 104 : tokens 232]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

outcry_[1] uprooted_[1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 2]

roach_[2]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

B. Families list

BNC-COCA-1,000 Families: [fams 90 : types 104 : tokens 232]

a_[10] about_[1] across_[1] after_[1] along_[1] although_[1] and_[10] art_[2]
as_[4] at_[2] be_[17] become_[1] begin_[1] between_[1] black_[1] break_[1]
by_[4] call_[1] carry_[1] child_[4] country_[1] day_[1] decide_[1] do_[1]
end_of_list_[1] even_[1] face_[1] family_[2] for_[2] force_[3] from_[3]
game_[1] general_[1] give_[1] govern_[1] have_[3] he_[14] high_[1]
hold_[1] home_[1] in_[9] into_[1] issue_[1] it_[2] late_[1] lead_[1] learn_[1]
left_[1] less_[1] light_[2] like_[1] live_[1] main_[1] man_[1] many_[2]
music_[1] nation_[1] not_[2] number_[10] of_[8] on_[1] one_[1] only_[2]
parent_[2] part_[1] particular_[1] past_[2] people_[1] place_[1] public_[1]
race_[1] should_[1] sing_[1] sister_[1] skin_[1] some_[1] song_[2] sorry_[1]
steal_[4] street_[1] support_[1] take_[2] that_[2] the_[13] there_[1] they_[6]
this_[4] through_[1] to_[5] until_[1] up_[2] walk_[1] way_[1] which_[1]
white_[2] with_[3] write_[1] wrong_[1] year_[2] young_[1]

BNC-COCA-2,000 Families: [fams 18 : types 19 : tokens 21]

anger_[1] attention_[1] avoid_[1] bridge_[1] century_[1] flame_[1]
generation_[2] hero_[1] lane_[1] mix_[1] practise_[2] react_[1] remove_[2]
respect_[1] society_[1] stream_[1] suffer_[1] tour_[1]

BNC-COCA-3,000 Families: [fams 12 : types 12 : tokens 13]

adopt_[1] album_[1] apology_[1] aspect_[1] athlete_[1] federal_[1] foster_[2]
guitar_[1] harbor_[1] immigrant_[1] media_[1] origin_[1]

BNC-COCA-4,000 Families: [fams 2 : types 2 : tokens 2]

debut_[1] reconcile_[1]

BNC-COCA-5,000 Families: [fams 2 : types 2 : tokens 2]

assimilate_[1] orphan_[1]

BNC-COCA-6,000 Families: [fams 1 : types 1 : tokens 1]

charcoal_[1]

BNC-COCA-7,000 Families: [fams 1 : types 2 : tokens 4]

aborigine_[4]

BNC-COCA-8,000 Families: [fams 2 : types 2 : tokens 2]

outcry_[1] uproot_[1]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 2]

roach [2]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.2.1.7 Native Americans in Business

7.2.1.7.1 Text only file

Native Americans in Business

Native American communities and individuals have been working their way into economic prosperity for quite some time now . While some of this has been due to the impacts of government programs set up to assist native business people , most of it has been an outgrowth of native ingenuity and innovation . Native Americans are in charge of some of the largest resource development companies , some of the largest restaurant chains , some of the largest casinos , and some very popular capital investment firms and financing companies all

over the country . While these natives represent a wide range of economic interests , one thing they share is the way their heritage has influenced the way they handle their business operations .

One example of a business man that made a big name for himself is David Anderson , the head of the " Famous Dave's " chain of restaurants . Dave Anderson is an Ojibwe native of Minnesota that grew up on and off reservations for most of his childhood . He was raised on traditional native values , and has applied those values throughout his business career . He helped to develop and create Rain forest Cafe , which has reached an immense level of profitability in its own rite . After attaining his Masters in Public Administration from Harvard , he worked for years to help struggling Native American businesses reach financial success . His dream was to open a barbeque restaurant , and in 1994 he was able to realize that dream . " Famous Dave's Barbeque " is now a major restaurant chain that can be found throughout the United States . In 2004 , Dave Anderson became the Assistant Secretary of the Interior for Indian Affairs . It was a position that gave him the opportunity to find ways to help native communities thrive under increasing economic pressures . He currently runs a non profit organization called the Life Skills Center for Leadership , whose mission is to provide leadership training and assistance to at risk youth in order to help them maximize their potential in life .

Native Americans have also become very influential in the financial world . Ho Chunk , Inc . is a very successful example of native ingenuity applied to finances and economic development . Started in 1994 with a mission to advance the economic interests of the Winnebago Indian Reservation in Nebraska . Ho Chunk , Inc . has taken a community with over 60% unemployment and turned it into one of the most financially stable communities in Native American society . They are essentially an investment firm that makes aggressive investment decisions that they think will benefit the Winnebago people , and they have found a great deal of success in this work . They currently employ over 1400 people and run some non profit organizations . The most recent of these non profits is the Ho Chunk Community Development Corporation . This organization focuses entirely on uplifting the quality of life on the reservation through housing, economic, and educational programs . To help community employment , Ho Chunk provides both necessary education support in the community and special consideration when applying for employment with Ho Chunk . The goal is to employ as many Winnebago natives as possible , injecting greater prosperity into the reservation .

There are a number of successful businesses run by native individuals and communities out there far too many to list . Some of these businesses are native corporations that control natural resources in pristine wilderness , others are simply individuals that had a great idea and ran with it . In all cases , successful native business people consistently note a strong tie to their heritage that pushes them to succeed . Native American business people are also statistically the most likely to contribute extensively to the communities they have come from , with most money being contributed to programs that enhance the education of Native American youth .

7.2.1.7.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

they've

3. Hyphenated words with hyphen removed:

non-profit, at-risk, Ho-Chunk

4. Compound words separated:

Businesspeople , businessman, rainforest

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Americans, David, Anderson, Harvard, Indian, Ho, Chunk, Winnebago, Nebraska, Dave, ojibwe, Minnesota,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Native (Americans), Rainforest Café, United States, Interior for (Indian) Affairs, Life Skills Center for Leadership, (Winnebago Indian) Reservation, (Ho-Chunk) Community Development Corporation

Note: Text related to illustrations have been included in the text analysis.

7.2.1.7.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Access - Native Americans in Busine (3.89 kb)

Words recategorized by user as 1k items (proper nouns etc): AMERICANS, DAVID, ANDERSON, HARVARD, INDIAN, HO-CHUNK , WINNEBAGO, NEBRASKA, DAVE, OJIBWE, MINNESOTA (total 30 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	149	191	480	77.29%
Function:	(261)	(42.03%)
Content:	(219)	(35.27%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(94)	(15.14%)

K2 Words (1001-2000):	15	19	25	4.03%
> Anglo-Sax:	(3)	(0.48%)
1k+2k	(81.32%)
AWL Words (academic):	26	35	52	8.37%
> Anglo-Sax:	(3)	(0.48%)
Off-List Words:	<u>2</u>	<u>34</u>	<u>64</u>	<u>10.31%</u>
	190+?	279	621	100%

Words in text (tokens):	621
Different words (types):	279
Type-token ratio:	0.45
Tokens per type:	2.23
Lex density (content words/total)	0.58

Pertaining to onlist only

Tokens:	557
Types:	245
Families:	190
Tokens per family:	2.93
Types per family:	1.29
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
77.29	77.29
4.03	81.32
8.37	89.69
10.31	100.00

A. AWL Tokens lists

AWL [26:35:52] administration assist assistance assistant attaining benefit communities communities communities communities communities community community community community consistently contribute contributed corporation corporations create economic economic economic economic economic economic enhance finances financial financial

financially financing focuses goal impacts individuals individuals individuals
 innovation investment investment investment major maximize potential
 range resource resources stable statistically traditional

Sublist 1

benefit consistently create economic economic economic economic
 economic economic finances financial financial financially financing
 individuals individuals individuals major

Sublist 2

administration assist assistance assistant communities communities
 communities communities communities community community community
 community focuses impacts investment investment investment potential
 range resource resources traditional

Sublist 3

contribute contributed corporation corporations maximize

Sublist 4

goal statistically

Sublist 5

stable

Sublist 6

enhance

Sublist 7

innovation

Sublist 9

attaining

B. AWL Types list

AWL types: [26:35:52] administration_[1] assist_[1] assistance_[1]
 assistant_[1] attaining_[1] benefit_[1] communities_[5] community_[4]
 consistently_[1] contribute_[1] contributed_[1] corporation_[1]
 corporations_[1] create_[1] economic_[6] enhance_[1] finances_[1]
 financial_[2] financially_[1] financing_[1] focuses_[1] goal_[1] impacts_[1]
 individuals_[3] innovation_[1] investment_[3] major_[1] maximize_[1]
 potential_[1] range_[1] resource_[1] resources_[1] stable_[1] statistically_[1]
 traditional_[1]

C. AWL Families list

AWL families: [26:35:52]

administer_[1] assist_[3] attain_[1] benefit_[1] community_[9] consist_[1]
 contribute_[2] corporate_[2] create_[1] economy_[6] enhance_[1]
 finance_[5] focus_[1] goal_[1] impact_[1] individual_[3] innovate_[1]
 invest_[3] major_[1] maximise_[1] potential_[1] range_[1] resource_[2]
 stable_[1] statistic_[1] tradition_[1]

AWL Fr non-cognate families: [families 3 : tokens 3] enhance_[1] goal_[1]
 range_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer
 information (integral, edit, proper, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	138 (61.88)	170 (60.50)	444 (71.38)	71.38
K-2 Words :	50 (22.42)	67 (23.84)	118 (18.97)	90.35
K-3 Words :	24 (10.76)	28 (9.96)	35 (5.63)	95.98
K-4 Words :	5 (2.24)	5 (1.78)	5 (0.80)	96.78
K-5 Words :	2 (0.90)	2 (0.71)	2 (0.32)	97.10
K-6 Words :	1 (0.45)	1 (0.36)	1 (0.16)	97.26
K-7 Words :	1 (0.45)	1 (0.36)	1 (0.16)	97.42
K-8 Words :	1 (0.45)	1 (0.36)	2 (0.32)	97.74
K-9 Words :				
K-10 Words :				
K-11 Words :	1 (0.45)	1 (0.36)	1 (0.16)	97.90
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				

K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	3 (1.07)	3 (0.48)	98.38
Total (unrounded)	223+?	281 (100)	622 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	622
Different words (types):	281
Type-token ratio:	0.45
Tokens per type:	2.21

Pertaining to onlist only

Tokens:	619
Types:	278
Families:	223
Tokens per Family :	2.78
Types per Family :	1.25

Current profile (token %)	
K-1 (71.38)	71.38
K-2 (18.97)	90.35
K-3 (5.63)	95.98
K-4 (0.80)	96.78
K-5 (0.32)	97.10
K-6 (0.16)	97.26
K-7 (0.16)	97.42
K-8 (0.32)	97.74
K-11 (0.16)	97.90
OFF (0.48)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 138 : types 170 : tokens 444]

a [14] able [1] after [1] all [2] also [2] americans [3] an [4] and [17] anderson [3] are [6] as [2] at [1] be [1] became [1] become [1] been [3] being [1] big [1] both [1] business [7] businesses [3] by [1] called [1] can [1] cases [1] center [1] charge [1] childhood [1] chunk [5] come [1] companies [2] consideration [1] control [1] country [1] dave [4] david [1] deal [1] dream [2] education [2] educational [1] employ [2] employment [2] far [1] find [1] for [7] forest [1] found [2] from [2] gave [1] government [1] great [2] greater [1] grew [1] had [1] handle [1] harvard [1] has [6] have [4] he [5] head [1] help [4] helped [1] him [1] himself [1] his [4] housing [1] idea [1] in [16] indian [2] interests [2] into [3] is [8] it [4] its [1] largest [3] leadership [2] level [1] life [3] list [1] made [1] major [1] makes [1] man [1] many [2] masters [1] minnesota [1] money [1] most [6] name [1] natural [1] nebraska [1] necessary [1] note [1] now [2] number [6] of [25] off [1] ojibwe [1] on [4] one [3] open [1] order [1] others [1] out [1] over [3] own [1] people [5] position [1] possible [1] programs [3] public [1] pushes [1] quite [1] rain [1] raised [1] ran [1] reach [1] reached [1] realize [1] recent [1] run [2] runs [1] set [1] share [1] simply [1] some [8] special [1] started [1] strong [1] support [1] taken [1] that [11] the [29] their [5] them [2] there [2] these [3] they [7] thing [1] think [1] this [3] those [1] through [1] throughout [2] tie [1] time [1] to [21] too [1] training [1] turned [1] under [1] unemployment [1] up [2] very [3] was [4] way [3] ways [1] when [1] which [1] while [2] whose [1] wide [1] will [1] winnebago [3] with [5] work [1] worked [1] working [1] world [1] years [1]

BNC-COCA-2,000 types: [fams 50 : types 67 : tokens 118]

advance [1] affairs [1] applied [2] applying [1] assist [1] assistance [1] assistant [1] benefit [1] capital [1] career [1] chain [2] chains [1] communities [5] community [4] contribute [1] contributed [1] create [1] currently [2] decisions [1] develop [1] development [3] due [1] economic [6] entirely [1] example [2] famous [2] finances [1] financial [2] financially [1] financing [1] firm [1] goal [1] increasing [1] individuals [3] influenced [1] influential [1] likely [1] mission [2] native [17] natives [2] non [3] operations [1] opportunity [1] organization [2] organizations [1] popular [1] pressures [1] provide [1] provides [1] quality [1] range [1] represent [1] reservation [3] reservations [1] restaurant [3] restaurants [1] risk [1] skills [1] society [1] stable [1] states [1] struggling [1] success [2] successful [3] traditional [1] united [1] values [2]

BNC-COCA-3,000 types: [fams 24 : types 28 : tokens 35]

administration [1] aggressive [1] consistently [1] corporation [1] corporations [1] enhance [1] essentially [1] extensively [1] firms [1] focuses [1] heritage [2] impacts [1] inc [2] injecting [1] innovation [1]

interior_[1] investment_[3] potential_[1] profit_[2] profitability_[1] profits_[1]
prosperity_[2] resource_[1] resources_[1] secretary_[1] statistically_[1]
succeed_[1] youth_[2]

BNC-COCA-4,000 types: [fams 5 : types 5 : tokens 5]

attaining_[1] cafe_[1] immense_[1] maximize_[1] thrive_[1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

rite_[1] wilderness_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

casinos_[1]

BNC-COCA-7,000 types: [fams 1 : types 1 : tokens 1]

pristine_[1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 2]

ingenuity_[2]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

outgrowth_[1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 3]

barbeque_[2] uplifting_[1]

B. Families list

BNC-COCA-1,000 Families: [fams 138 : types 170 : tokens 444]

a_[18] able_[1] after_[1] all_[2] also_[2] and_[17] anderson_[1] as_[2] at_[1]
be_[23] become_[2] big_[1] both_[1] business_[10] by_[1] call_[1] can_[1]
case_[1] centre_[1] charge_[1] child_[1] chunk_[2] come_[1] company_[2]
consider_[1] control_[1] country_[1] dave_[1] david_[3] deal_[1] dream_[2]
educate_[3] employ_[5] end_of_list_[1] far_[1] find_[3] for_[7] forest_[1]
from_[2] give_[1] govern_[1] great_[3] grow_[1] handle_[1] harvard_[3]
have_[11] he_[11] head_[1] help_[5] house_[1] idea_[1] in_[16] indian_[1]
interest_[2] into_[3] it_[5] large_[3] lead_[2] level_[1] life_[3] list_[1]
major_[1] make_[2] man_[1] many_[2] master_[1] minnesota_[1] money_[1]
most_[6] name_[1] nature_[1] nebraska_[3] necessary_[1] note_[1] now_[2]
number_[6] of_[25] off_[1] ojbwe_[4] on_[4] one_[3] open_[1] order_[1]
other_[1] out_[1] over_[3] own_[1] people_[5] position_[1] possible_[1]
programme_[3] public_[1] push_[1] quite_[1] rain_[1] raise_[1] reach_[2]
realise_[1] recent_[1] run_[4] set_[1] share_[1] simple_[1] some_[8]
special_[1] start_[1] strong_[1] support_[1] take_[1] that_[12] the_[29]
there_[2] they_[14] thing_[1] think_[1] this_[6] through_[3] tie_[1] time_[1]
to_[21] too_[1] train_[1] turn_[1] under_[1] up_[2] very_[3] way_[4] when_[1]
which_[1] while_[2] who_[1] wide_[1] will_[1] winnebago_[5] with_[5]
work_[3] world_[1] year_[1]

BNC-COCA-2,000 Families: [fams 50 : types 67 : tokens 118]

advance_[1] affair_[1] apply_[3] assist_[3] benefit_[1] capital_[1] career_[1]
chain_[3] community_[9] contribute_[2] create_[1] current_[2] decision_[1]
develop_[4] due_[1] economy_[6] entire_[1] example_[2] famous_[2]

finance_[5] firm_[1] goal_[1] increase_[1] individual_[3] influence_[2]
likely_[1] mission_[2] native_[19] non_[3] operate_[1] opportunity_[1]
organize_[3] popular_[1] pressure_[1] provide_[2] quality_[1] range_[1]
represent_[1] reserve_[4] restaurant_[4] risk_[1] skill_[1] society_[1]
stable_[1] states_[1] struggle_[1] success_[5] tradition_[1] unite_[1]
value_[2]

BNC-COCA-3,000 Families: [fams 24 : types 28 : tokens 35]

administration_[1] aggressive_[1] consistent_[1] corporate_[2] enhance_[1]
essential_[1] extensive_[1] firms_[1] focus_[1] heritage_[2] impact_[1]
incorporate_[2] inject_[1] innovate_[1] interior_[1] invest_[3] potential_[1]
profit_[4] prosper_[2] resource_[2] secretary_[1] statistic_[1] succeed_[1]
youth_[2]

BNC-COCA-4,000 Families: [fams 5 : types 5 : tokens 5]

attain_[1] cafe_[1] immense_[1] maximise_[1] thrive_[1]

BNC-COCA-5,000 Families: [fams 2 : types 2 : tokens 2]

rite_[1] wilderness_[1]

BNC-COCA-6,000 Families: [fams 1 : types 1 : tokens 1]

casino_[1]

BNC-COCA-7,000 Families: [fams 1 : types 1 : tokens 1]

pristine_[1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 2]

ingenuity_[2]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens]

outgrowth_[1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 3]

barbeque_[2] uplifting_[1]

7.2.2 Stunt

7.2.2.1 English as a World Language

7.2.2.1.1 Text only file

English is a language most of us are exposed to every day . Is this true for you ? On average how much time to you think you spend listening to , reading , writing or speaking English ?

English as a World Language

As John Donne says , no man is an island . In our modern world it is impossible to isolate ourselves from any kind of interaction with others . Most of us meet and communicate with other people all the time .

For thousands of years , people have been travelling . If you travel a short distance you can usually manage by using your own language , but as soon as you get further from home you need another means of communication . When we travel today we usually speak English . In many countries , it is an official language , but we can also communicate with people in

countries where English does not have this status . According to the famous linguist David Crystal , English is spoken by 1.2 - 1.5 billion people around the world . Only about 350 million of these have English as their mother tongue .

When a language is used for communication by two people who are not native speakers, we call it a lingua franca . The world's most important lingua franca today is English . Why has the English language become so important ?

From the time of the reign of Queen Elizabeth I and for the next 400 years , the British travelled around the world building their empire . As they established colonies on every continent , the English language was spread and adopted by people in all corners of the world . This is why the English language is still an official language in 52 countries today . During the reign of Queen Victoria , in the nineteenth century , the British Empire experienced its golden age . About 25 % of the world's population and about 25 % of all land territory belonged to the Empire . This gave

In the first half of the twentieth century, Britain suffered great losses in the two world wars . These losses , coupled with major political turmoil in Africa and the Indian sub continent , meant that Britain lost its position as the world's leading power . However , another English speaking nation , the USA , became a super power , and this ensured the continuing international importance of the English language . Because of the roles that these two countries have played in world history , English has become important not only for native speakers , but for millions of other people .

Most people in the western world are exposed to it every day ; in today's Norway there is hardly a job you can do a subject you can study or an Internet page you can

" English is a language which has repeatedly found itself in the right place at the right time "

- David Crystal 1941 - , English linguist .

visit that does not require a minimum basic knowledge of English . We are surrounded by it . From the time English became a world language , it has developed in different directions . The language is in constant change and everyone who speaks it participates in this development . In different parts of the world , there are different variants of spelling , pronunciation , grammar and vocabulary .

The English language has become the world's most important lingua franca . This gives the language a dominant position in trade , politics and culture . At the same time , it also gives all English speakers , native and non native , a share in the language .

7.2.2.1.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

English-speaking, non-native, sub-continent

4. Compound words separated:

superpower

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

John, Donne, English, David, Crystal, British, Elizabeth, Victoria, USA, Norway, Indian, Africa, Britain, Internet

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Queen (Elizabeth) I, Queen (Victoria), (British) Empire,

Note: Text related to illustrations have been included in the text analysis.

7.2.2.1.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Stunt English as a World Language (3.35 kb)

Words recategorized by user as 1k items (proper nouns etc): JOHN, DONNE, ENGLISH, DAVID, CRYSTAL, BRITISH, ELIZABETH, VICTORIA, USA, NORWAY, INDIAN, AFRICA, BRITAIN, INTERNET (total 37 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	149	180	496	87.63%
Function:	(257)	(45.41%)
Content:	(239)	(42.23%)
> Anglo-Sax	(125)	(22.08%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	11	11	12	2.12%
> Anglo-Sax:	(3)	(0.53%)
1k+2k		(89.75%)
AWL Words (academic):	18	19	22	3.89%
> Anglo-Sax:	()	(0.00%)
Off-List Words:	<u>2</u>	<u>25</u>	<u>36</u>	<u>6.36%</u>
	178+?	235	566	100%

Words in text (tokens): 566

Different words (types):	235
Type-token ratio:	0.42
Tokens per type:	2.41
Lex density (content words/total)	0.55

Pertaining to onlist only

Tokens:	530
Types:	210
Families:	178
Tokens per family:	2.98
Types per family:	1.18
Anglo-Sax Index: <small>(A-Sax tokens + functors / onlist tokens)</small>	%
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
87.63	87.63
2.12	89.75
3.89	93.64
6.36	100.00

A. AWL Tokens lists

AWL [18:19:22] communicate communicate communication communication constant coupled culture dominant ensured established exposed exposed interaction isolate job major minimum participates require roles status variants

Sublist 1

established major require roles variants

Sublist 2

culture participates

Sublist 3

constant dominant ensured interaction

Sublist 4

communicate communicate communication communication job status

Sublist 5

exposed exposed

Sublist 6

minimum

Sublist 7

coupled isolate

B. AWL Types list

AWL types: [18:19:22] communicate_[2] communication_[2] constant_[1] coupled_[1] culture_[1] dominant_[1] ensured_[1] established_[1] exposed_[2] interaction_[1] isolate_[1] job_[1] major_[1] minimum_[1] participates_[1] require_[1] roles_[1] status_[1] variants_[1]

C. AWL Families list

AWL families: [18:19:22]

communicate_[4] constant_[1] couple_[1] culture_[1] dominate_[1] ensure_[1] establish_[1] expose_[2] interact_[1] isolate_[1] job_[1] major_[1] minimum_[1] participate_[1] require_[1] role_[1] status_[1] vary_[1]

AWL Fr non-cognate families: [families : tokens]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	147 (72.77)	174 (73.73)	472 (83.39)	83.39
K-2 Words :	35 (17.33)	37 (15.68)	61 (10.78)	94.17
K-3 Words :	14 (6.93)	15 (6.36)	18 (3.18)	97.35
K-4 Words :	1 (0.50)	1 (0.42)	2 (0.35)	97.70
K-5 Words :	2 (0.99)	2 (0.85)	2 (0.35)	98.05
K-6 Words :	1 (0.50)	1 (0.42)	1 (0.18)	98.23
K-7 Words :				
K-8 Words :	1 (0.50)	1 (0.42)	2 (0.35)	98.58

K-9 Words :				
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	1 (0.50)	1 (0.42)	3 (0.53)	99.11
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	3 (1.27)	4 (0.71)	99.82
Total (unrounded) 202+?		236 (100)	566 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	566
Different words (types):	236
Type-token ratio:	0.42
Tokens per type:	2.40

Pertaining to onlist only

Tokens:	562
Types:	233
Families:	202
Tokens per Family :	2.78
Types per Family :	1.15

Current profile (token %)	
K-1 (83.39)	83.39
K-2 (10.78)	94.17
K-3 (3.18)	97.35
K-4 (0.35)	97.70
K-5 (0.35)	98.05
K-6 (0.18)	98.23
K-8 (0.35)	98.58
K-15 (0.53)	99.11
OFF (0.71)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 124 : types 146 : tokens 437]

a_[13] about_[3] age_[1] all_[4] also_[2] an_[4] and_[10] another_[2] any_[1] are_[5] around_[2] as_[7] at_[2] basic_[1] became_[2] because_[1] become_[3] been_[1] billion_[1] building_[1] but_[3] by_[5] call_[1] can_[5] change_[1] continuing_[1] corners_[1] countries_[4] **coupled**_[1] day_[2] different_[3] do_[1] does_[2] during_[1] every_[3] everyone_[1] **experienced**_[1] first_[1] for_[6] found_[1] from_[4] further_[1] gave_[1] get_[1] gives_[2] golden_[1] great_[1] half_[1] hardly_[1] has_[5] have_[4] history_[1] home_[1] how_[1] however_[1] i_[1] if_[1] important_[4] impossible_[1] in_[19] internet_[1] is_[13] island_[1] it_[8] its_[2] itself_[1] job_[1] kind_[1] land_[1] leading_[1] listening_[1] lost_[1] major_[1] man_[1] manage_[1] many_[1] **means**_[1] meant_[1] meet_[1] million_[1] millions_[1] most_[5] mother_[1] much_[1] nation_[1] need_[1] next_[1] nineteenth_[1] no_[1] not_[4] number_[10] of_[19] on_[2] only_[2] or_[2] other_[2] others_[1] our_[1] ourselves_[1] own_[1] page_[1] parts_[1] people_[8] place_[1] played_[1] position_[2] power_[2] queen_[2] reading_[1] right_[2] same_[1] says_[1] share_[1] short_[1] so_[1] soon_[1] speak_[1] speakers_[3] speaking_[2] speaks_[1] spend_[1] spoken_[1] still_[1] study_[1] subject_[1] that_[3] the_[38] their_[2] there_[2] these_[3] they_[1] think_[1] this_[7] thousands_[1] time_[6] to_[7] today_[4] travel_[2] travelled_[1] travelling_[1] true_[1] twentieth_[1] two_[3] us_[2] used_[1] using_[1] usually_[2] visit_[1] wars_[1] was_[1] we_[5] when_[2] where_[1] which_[1] who_[2] why_[2] with_[4] world_[14] writing_[1] years_[2] you_[10] your_[1]

BNC-COCA-2,000 types: [fams 35 : types 37 : tokens 61]

according_[1] average_[1] belonged_[1] century_[2] constant_[1] culture_[1] developed_[1] development_[1] directions_[1] distance_[1] empire_[3]

established_[1] exposed_[2] famous_[1] knowledge_[1] language_[16]
losses_[2] modern_[1] native_[4] non_[1] official_[2] political_[1] politics_[1]
population_[1] pronunciation_[1] repeatedly_[1] require_[1] roles_[1]
spelling_[1] spread_[1] suffered_[1] super_[1] surrounded_[1] tongue_[1]
trade_[1] variants_[1] western_[1]

BNC-COCA-3,000 types: [fams 15 : types 16 : tokens 20]

adopted_[1] colonies_[1] communicate_[2] communication_[2] continent_[2]
crystal_[2] dominant_[1] ensured_[1] importance_[1] interaction_[1]
international_[1] isolate_[1] minimum_[1] participates_[1] status_[1]
territory_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 2]

reign_[2]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

grammar_[1] vocabulary_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

harlot_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 2]

linguist_[2]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 3]

lingua_[3]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 4]

franca_[3] sub_[1]

B. Families list

BNC-COCA-1,000 Families: [fams 124 : types 146 : tokens 437]

a_[17] about_[3] age_[1] all_[4] also_[2] and_[10] another_[2] any_[1]
around_[2] as_[7] at_[2] basic_[1] be_[20] because_[1] become_[5]
billion_[1] build_[1] but_[3] by_[5] call_[1] can_[5] change_[1] continue_[1]
corner_[1] country_[4] couple_[1] day_[2] different_[3] do_[3] during_[1]
every_[4] experience_[1] find_[1] first_[1] for_[6] from_[4] further_[1] get_[1]
give_[3] gold_[1] great_[1] half_[1] hardly_[1] have_[9] history_[1] home_[1]
how_[1] however_[1] i_[1] if_[1] important_[4] in_[19] internet_[1] island_[1]
it_[11] job_[1] kind_[1] land_[1] lead_[1] listen_[1] lose_[1] major_[1]
man_[1] manage_[1] many_[1] mean_[2] meet_[1] million_[2] most_[5]
mother_[1] much_[1] nation_[1] need_[1] next_[1] nine_[1] no_[1] not_[4]
number_[10] of_[19] on_[2] only_[2] or_[2] other_[3] own_[1] page_[1]
part_[1] people_[8] place_[1] play_[1] position_[2] possible_[1] power_[2]
queen_[2] read_[1] right_[2] same_[1] say_[1] share_[1] short_[1] so_[1]
soon_[1] speak_[8] spend_[1] still_[1] study_[1] subject_[1] that_[3] the_[38]

there_2] they_3] think_1] this_10] thousand_1] time_6] to_7] today_4]
travel_4] true_1] twenty_1] two_3] use_2] usual_2] visit_1] war_1]
we_9] when_2] where_1] which_1] who_2] why_2] with_4] world_14]
write_1] year_2] you_11]

BNC-COCA-2,000 Families: [fams 35 : types 37 : tokens 61]

according_1] average_1] belong_1] century_2] constant_1] culture_1]
develop_2] direction_1] distance_1] empire_3] establish_1] expose_2]
famous_1] knowledge_1] language_16] loss_2] modern_1] native_4]
non_1] official_2] politics_2] population_1] pronounce_1] repeat_1]
require_1] role_1] spell_1] spread_1] suffer_1] super_1] surround_1]
tongue_1] trade_1] vary_1] western_1]

BNC-COCA-3,000 Families: [fams 15 : types 16 : tokens

adopt_1] colony_1] communicate_4] continent_2] crystal_2]
dominant_1] ensure_1] importance_1] interact_1] international_1]
isolate_1] minimum_1] participate_1] status_1] territory_1]

BNC-COCA-4,000 Families: [fams 1 : types 1 : tokens 2]

reign_2]

BNC-COCA-5,000 Families: [fams 2 : types 2 : tokens 2]

grammar_1] vocabulary_1]

BNC-COCA-6,000 Families: [fams 1 : types 1 : tokens 1]

turmoil_1]

BNC-COCA-7,000 Families: [fams : types : tokens]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 2]

linguist_2]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens

lingua_3]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 4]

franca_3] sub_1]

7.2.2.2 *British vs. American English*

7.2.2.2.1 Text only file

Starting Point

George Bernard Shaw said that the Americans and the British were "divided by a common language". What do you think he meant? Can you think of any differences?

British versus American English

British and American English, as well as all other "Englishes", come from the same source, the English of the British Isles. However, time and distance have changed several aspects of the language spoken by the colonists and imperialists of the 17th century. As students of

English you must make a conscious choice concerning which of these two forms you will use and you should try to be consistent . In order to do that , you need to know about some of the differences .

The first difference is evident to the ear . Without having to analyze what people say , we can immediately hear the difference between Americans and Britons in the way they speak and the words they choose . Some people even feel that the British sound more intelligent and the Americans more gregarious , but this depends on the speaker and the listener . Since both of these English forms also have many variations , like Scottish or a southern accent, let us narrow pronunciation down to Received Pronunciation for British English and General American for American English . In Received Pronunciation the vowel sounds are often rounder than in General American , while the consonant which is pronounced at the end of a word in General American is not pronounced in Received Pronunciation .

Try pronouncing these words in Received Pronunciation remember rounder vowels , and then in General American flatter vowels

Awful Dance Rather Laugh

Then practice these words with no in Received Pronunciation and a pronounced in General American :

Car Star Pillar

The other pronunciation difference is found in words of several syllables , such as " advertisement " . In Received Pronunciation the emphasis is on the second syllable :

advertisement, while in General American the emphasis is on , while in General American the emphasis is on

English

" There even are places where English completely disappears .

In America , they have not used it for years ! Why can not the English teach their children how to speak ? "

Alan Jay Lerner 1918 - 1986 , American song writer . " Why Can not the English ? " 1956 song .

the third : The clue here is to use your dictionaries , the ones on line and the ones in book form . All dictionaries show pronunciation , but many have different pronunciation keys . Find one that you like and use it often .

The second difference between British English and American English is in spelling and punctuation . This is the area you need to pay attention to when writing . The most important punctuation difference is in the use of what are known in British English as inverted commas and in American English as quotation marks . These are used in quotes or around titles . In British English you use one inverted comma To be or not to be , while in American English you use two " ... that is the question . " This can get a bit tricky .

Sometimes you need a set of quotation marks within quotation marks . In that case , you use the opposite form . For example , Kristin once said , In

my opinion , " The Road Not Taken " is the best poem ever written . To make matters even worse , British English often uses [""] around quotes but [""] around titles .

When it comes to the most common spelling differences , these can be narrowed down to five :

1. Some words that end in in American English end in in British English

e.g. theater American English theatre British English .

2. Some words that end in in American English end in in British English

5. In British English , the at the end of a word is doubled when a new ending is added . In American English the is only doubled if the stress is originally on the second syllable :

e.g. rebel - rebelled American English rebel - rebelled British English travel - traveler American English travel - traveller British English

The third and last difference between American English and British English is in usage . That is , words can mean different things and the same thing can have different words . This last category is very important , and can easily lead to embarrassing situations . Many British English speakers studying in the US have asked to borrow a " rubber " from another student .

The American student is understandably shocked . In the US , " rubber " is slang for " condom " : while in British English it only means an instrument to erase pencil marks , an eraser . " Bloody " is another example . In British English this is a profanity ; in American English it just means something full of blood .

Here are some examples of vocabulary differences :

American English	British English	Norwegian
------------------	-----------------	-----------

elevator	lift	
----------	------	--

sidewalk	pavement	road trip	car journey	popsicle	ice lolly
----------	----------	-----------	-------------	----------	-----------

stand in line	queue				
---------------	-------	--	--	--	--

bath room	water closet				
-----------	--------------	--	--	--	--

e.g. monolog American English catalog American English

monologue British English catalogue British English

eraser rubber

truck lorry

3. Some words that end in in American English end in in British English.

e.g. color American English colour British English

4. Verbs that end in in American English end in in British English

e.g. analyze American English analyse British English

VOCABULARY

To sum up both varieties of English are English, but there are a number of differences. Unfortunately, learners of English must be consistent users of either American English or British English, especially when writing. Fortunately, there are those tools called "dictionaries". They show differences in spelling, punctuation, usage and pronunciation; some even let you hear the pronunciation differences. In addition to dictionaries, be sure to use computer spell checks. These allow you to choose your type of English and let you know if you are mixing them. Maybe Shaw was right? Maybe the British and the Americans are "divided by a common language"?

Other former colonies, like New Zealand Australia and Canada have kept the British spelling but have their own distinct accents and some particular vocabulary.

Verb Conjugations

English: I walk, You walk, He she it walks we walk you walk they walk

We are often told that English is a simple language, and therefore easy to learn. However, this is not always the case. Look at the verb tables above which show present tense forms. We see that here Norwegian is simpler as we do not conjugate our verbs according to number or person. In English you have to add an to the third person singular. Italian is even more complicated with all persons in both the singular and the plural having separate forms. As a result they often leave out the pronoun since it is unnecessary. The form of the verb tells you who the subject is. They would for instance say I am Kristin instead of "Io sono Kristin".

7.2.2.2.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions and abriviations that are written out:

haven't, can't (2)

3rd - written out as "third"

vs. – written out as versus

RP (Received Pronunciation), GA (General American), BE (British English) and AE (American English), WC (water closet)

3. Hyphenated words with hyphen removed:

4. Compound words separated:

Songwriter, online, bookform, bathroom

5. Words (groups of letters) removed from the text analysis:

Word parts that were removed: -ter, -tre, -l (2), -ize/yzze, -ise/yse, -or, -our, -s, -og, -ogue

In text translations to Norwegian were removed: stå i kue, ispinne, viskelær, lastebil, fortau, heis, biltur

«r's»

6. Proper nouns:

English, British, American, Americans, Englishes, Britons, Scottish, George, Bernard, Shaw, Alan, Jay, Lerner, New Zealand, Kristin, Norwegian, Australia, Canada, Italian

Take note: The words outside of brackets have not been placed on the list of proper nouns.

New (Zealand)

Note: Text related to illustrations have been included in the text analysis.

7.2.2.2.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: British vs. American English Stunt (6.82 kb)

Words recategorized by user as 1k items (proper nouns etc): ENGLISH, BRITISH, AMERICAN, AMERICANS, ENGLISHES, BRITONS, SCOTTISH, GEORGE, BERNARD, SHAW, ALAN, JAY, LERNER, ZEALAND, KRISTIN, NORWEGIAN, AUSTRALIA, CANADA, ITALIAN (total 142 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	205	260	863	78.17%
Function:	(478)	(43.30%)
Content:	(385)	(34.87%)
> Anglo-Sax	(227)	(20.56%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	37	42	59	5.34%
> Anglo-Sax:	(19)	(1.72%)
1k+2k		(83.51%)
AWL Words (academic):	17	19	26	2.36%
> Anglo-Sax:	()	(0.00%)
Off-List Words:	?	<u>69</u>	<u>156</u>	<u>14.13%</u>
	259+?	390	1104	100%
Words in text (tokens):		1104		
Different words (types):		390		
Type-token ratio:		0.35		
Tokens per type:		2.83		
Lex density (content words/total)		0.57		

Pertaining to onlist only

Tokens:	948
Types:	321
Families:	259
Tokens per family:	3.66
Types per family:	1.24
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul
78.17	78.17
5.34	83.51
2.36	85.87
14.13	100.00

A. AWL Tokens lists

AWL [17:19:26] analyse analyze analyze area aspects category computer consistent consistent distinct emphasis emphasis emphasis evident instance intelligent quotation quotation quotation quotes quotes source stress sum tense variations

Sublist 1

analyse analyze analyze area consistent consistent evident source variations

Sublist 2

aspects category computer distinct

Sublist 3

emphasis emphasis emphasis instance

Sublist 4

stress sum

Sublist 6

intelligent

Sublist 7

quotation quotation quotation quotes quotes

Sublist 8

tense

B. AWL Types list

AWL types: [17:19:26] analyse_[1] analyze_[2] area_[1] aspects_[1] category_[1] computer_[1] consistent_[2] distinct_[1] emphasis_[3] evident_[1] instance_[1] intelligent_[1] quotation_[3] quotes_[2] source_[1] stress_[1] sum_[1] tense_[1] variations_[1]

C. AWL Families list

AWL families: [17:19:26]

analyse_[3] area_[1] aspect_[1] category_[1] compute_[1] consist_[2] distinct_[1] emphasis_[3] evident_[1] instance_[1] intelligence_[1] quote_[5] source_[1] stress_[1] sum_[1] tense_[1] vary_[1]

AWL Fr non-cognate families: [families : tokens]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: British vs. American English - Stunt (6,993 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): english, british, american, americans, englishes, britons, scottish, george, bernard, shaw, alan, jay, lerner, zealand, kristin, norwegian, australia, canada, italian end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level Families (%) Types (%) Tokens (%) Cumul. token %

K-1 Words :	225 (70.98)	277 (70.13)	936 (83.80)	83.80
K-2 Words :	43 (13.56)	51 (12.91)	83 (7.43)	91.23
K-3 Words :	18 (5.68)	22 (5.57)	28 (2.51)	93.74
K-4 Words :	5 (1.58)	5 (1.27)	10 (0.90)	94.64
K-5 Words :	8 (2.52)	10 (2.53)	18 (1.61)	96.25
K-6 Words :	6 (1.89)	7 (1.77)	11 (0.98)	97.23
K-7 Words :	1 (0.32)	2 (0.51)	3 (0.27)	97.50
K-8 Words :	5 (1.58)	5 (1.27)	5 (0.45)	97.95
K-9 Words :	2 (0.63)	3 (0.76)	3 (0.27)	98.22
K-10 Words :	1 (0.32)	1 (0.25)	1 (0.09)	98.31
K-11 Words :	1 (0.32)	2 (0.51)	2 (0.18)	98.49
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	1 (0.32)	1 (0.25)	1 (0.09)	98.58
K-16 Words :				
K-17 Words :	1 (0.32)	1 (0.25)	1 (0.09)	98.67
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	4 (1.01)	4 (0.36)	99.03
Total (unrounded)	317+?	395 (100)	1117 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1117
Different words (types):	395
Type-token ratio:	0.35

Tokens per type: 2.83

Pertaining to onlist only

Tokens:	1113
Types:	391
Families:	317
Tokens per Family :	3.51
Types per Family :	1.23

Current profile (token %)	
K-1 (83.80)	83.80
K-2 (7.43)	91.23
K-3 (2.51)	93.74
K-4 (0.90)	94.64
K-5 (1.61)	96.25
K-6 (0.98)	97.23
K-7 (0.27)	97.50
K-8 (0.45)	97.95
K-9 (0.27)	98.22
K-10 (0.09)	98.31
K-11 (0.18)	98.49
K-15 (0.09)	98.58
K-17 (0.09)	98.67
OFF (0.36)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 184 : types 220 : tokens 794]

a_[15] about_[1] above_[1] add_[1] added_[1] addition_[1] advertisement_[2] all_[3] allow_[1] also_[1] always_[1] am_[1] an_[3] and_[28] another_[2] any_[1] are_[11] area_[1] around_[3] as_[8] asked_[1] at_[3] awful_[1] bath_[1] be_[6] best_[1] between_[3] bit_[1] blood_[1] bloody_[1] book_[1] both_[3] but_[5] by_[3] called_[1] can_[9] car_[2] case_[2] changed_[1] checks_[1] children_[1] choice_[1] choose_[2] color_[1] colour_[1] come_[1] comes_[1] completely_[1] computer_[1] concerning_[1] dance_[1] depends_[1] difference_[6] differences_[7] different_[3] do_[3] doubled_[2] down_[2] ear_[1] easily_[1] easy_[1] either_[1] end_[10] ending_[1] especially_[1] even_[5] ever_[1] feel_[1] find_[1] first_[1] five_[1] for_[6] form_[3] forms_[4] fortunately_[1] found_[1] from_[2] full_[1] general_[7] get_[1] have_[9] having_[2] he_[2] hear_[2]

here [3] how [1] however [2] i [2] ice [1] if [2] important [2] in [54]
instead [1] is [30] it [7] italian [1] just [1] kept [1] keys [1] know [2]
known [1] last [2] laugh [1] lead [1] learn [1] learners [1] leave [1] let [3]
lift [1] like [3] line [2] listener [1] look [1] make [2] many [3] marks [4]
matters [1] maybe [2] mean [1] means [2] meant [1] more [3] most [2]
must [2] my [1] need [3] new [2] no [1] not [8] number [11] of [22]
often [5] on [6] once [1] one [2] ones [2] only [2] or [5] order [1]
other [3] our [1] out [1] own [1] particular [1] pay [1] people [2]
person [2] persons [1] places [1] point [1] present [1] question [1]
rather [1] remember [1] right [1] road [2] room [1] rounder [2] said [2]
same [2] say [2] second [3] see [1] set [1] several [2] she [1] should [1]
show [3] simple [1] simpler [1] since [2] situations [1] some [8]
something [1] sometimes [1] song [2] sound [1] sounds [1] speak [2]
speaker [1] speakers [1] spoken [1] stand [1] star [1] starting [1]
student [2] students [1] studying [1] subject [1] such [1] sure [1]
tables [1] taken [1] teach [1] tells [1] than [1] that [13] the [64] their [2]
them [1] then [2] there [3] these [7] they [7] thing [1] things [1] think [2]
third [3] this [6] those [1] time [1] to [26] told [1] travel [2] traveler [1]
traveller [1] trip [1] try [2] two [2] type [1] understandably [1]
unfortunately [1] unnecessary [1] up [1] us [3] use [8] used [2] users [1]
uses [1] very [1] walk [5] walks [1] was [1] water [1] way [1] we [5]
well [1] were [1] what [3] when [4] where [1] which [3] while [5] who [1]
why [2] will [1] with [2] within [1] without [1] word [2] words [9] worse [1]
would [1] writer [1] writing [2] written [1] years [1] you [20] your [2]

BNC-COCA-2,000 types: [fams 43 : types 50 : tokens 83]

accent [1] accents [1] according [1] attention [1] borrow [1] century [1]
clue [1] common [3] complicated [1] conscious [1] disappears [1]
distance [1] divided [2] embarrassing [1] example [2] examples [1]
immediately [1] instance [1] instrument [1] journey [1] language [4]
mixing [1] narrow [1] narrowed [1] opinion [1] opposite [1] originally [1]
poem [1] practice [1] pronounced [3] pronouncing [1] pronunciation [12]
quotation [3] quotes [2] received [6] result [1] separate [1] shocked [1]
southern [1] spell [1] spelling [4] stress [1] tense [1] theater [1]
theatre [1] therefore [1] titles [2] tools [1] tricky [1] truck [1] variations [1]

BNC-COCA-3,000 types: [fams 3 : types 3 : tokens 3]

analyse [1] analyze [2] aspects [1] catalog [1] catalogue [1] category [1]
colonies [1] colonists [1] consistent [2] distinct [1] elevator [1]
emphasis [3] evidenti [1] former [1] imperialists [1] pavement [1] rebel [2]
rebelled [2] source [1] sum [1] varieties [1] versus [1]

BNC-COCA-4,000 types: [fams 5 : types 5 : tokens 10]

dictionaries [4] intelligent [1] pencil [1] plural [1] rubber [3]

BNC-COCA-5,000 types: [fams 8 : types 9 : tokens 18]

condom [1] erase [1] eraser [2] flatter [1] pillar [1] singular [2] usage [2]
verb [3] verbs [2] vocabulary [3]

BNC-COCA-6,000 types: [fams 6 : types 6 : tokens 11]

closet [1] inverted [2] isles [1] punctuation [3] queue [1] syllable [2]
syllables [1]

BNC-COCA-7,000 types: [fams 1 : types 2 : tokens 3]

vowel [1] vowels [2]

BNC-COCA-8,000 types: [fams 6 : types 6 : tokens 6]

consonant [1] jay [1] lorry [1] monologue [1] pronoun [1] slang [1]

BNC-COCA-9,000 types: [fams 2 : types 2 : tokens 3]

comma [1] commas [1] profanity [1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

gregarious [1]

BNC-COCA-11,000 types: [fams 1 : types 2 : tokens 2]

conjugate [1] conjugations [1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens]

lolly [1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens 1]

popsicle_[1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

monolog_[1] sidewalk_[1]

B. Families list

BNC-COCA-1,000 Families: [fams 184 : types 220 : tokens 794]

a_[18] about_[1] above_[1] add_[3] advertise_[2] all_[3] allow_[1] also_[1]
always_[1] and_[28] another_[2] any_[1] area_[1] around_[3] as_[8] ask_[1]
at_[3] awful_[1] bath_[1] be_[50] best_[1] between_[3] bit_[1] blood_[2]
book_[1] both_[3] but_[5] by_[3] call_[1] can_[9] car_[2] case_[2]
change_[1] check_[1] child_[1] choice_[1] choose_[2] colour_[2] come_[2]
complete_[1] computer_[1] concern_[1] dance_[1] depend_[1]
difference_[13] different_[3] do_[3] double_[2] down_[2] ear_[1] easy_[2]
either_[1] end_[11] end_of_list_[1] especially_[1] even_[5] ever_[1] feel_[1]
find_[2] first_[1] five_[1] for_[6] form_[7] fortunate_[2] from_[2] full_[1]
general_[7] get_[1] have_[11] he_[2] hear_[2] here_[3] how_[1] however_[2]
i_[3] ice_[1] if_[2] important_[2] in_[54] instead_[1] it_[7] just_[1] keep_[1]
key_[1] know_[3] last_[2] laugh_[1] lead_[1] learn_[2] leave_[1] let_[3]
lift_[1] like_[3] line_[2] listen_[1] look_[1] make_[2] many_[3] mark_[4]
matter_[1] maybe_[2] mean_[4] more_[3] most_[2] must_[2] necessary_[1]

need_[3] new_[2] no_[1] not_[8] number_[11] of_[22] often_[5] on_[6]
once_[1] one_[4] only_[2] or_[5] order_[1] other_[3] out_[1] own_[1]
particular_[1] pay_[1] people_[2] person_[3] place_[1] point_[1] present_[1]
question_[1] rather_[1] remember_[1] right_[1] road_[2] room_[1] round_[2]
same_[2] say_[4] second_[3] see_[1] set_[1] several_[2] she_[1] should_[1]
show_[3] simple_[2] since_[2] situation_[1] some_[10] song_[2] sound_[2]
speak_[5] stand_[1] star_[1] start_[1] student_[3] study_[1] subject_[1]
such_[1] sure_[1] table_[1] take_[1] teach_[1] tell_[2] than_[1] that_[14]
the_[64] then_[2] there_[3] they_[10] thing_[2] think_[2] this_[13] three_[3]
time_[1] to_[26] travel_[4] trip_[1] try_[2] two_[2] type_[1] understand_[1]
up_[1] use_[12] very_[1] walk_[6] water_[1] way_[1] we_[9] well_[1]
what_[3] when_[4] where_[1] which_[3] while_[5] who_[1] why_[2] will_[1]
with_[2] within_[1] without_[1] word_[11] worse_[1] would_[1] write_[4]
year_[1] you_[22]

BNC-COCA-2,000 Families: [fams 43 : types 50 : tokens 83]

accent_[2] according_[1] attention_[1] borrow_[1] century_[1] clue_[1]
common_[3] complicate_[1] conscious_[1] disappear_[1] distance_[1]
divide_[2] embarrass_[1] example_[3] immediate_[1] instance_[1]
instrument_[1] journey_[1] language_[4] mix_[1] narrow_[2] opinion_[1]
opposite_[1] original_[1] poem_[1] practise_[1] pronounce_[16] quote_[5]
receive_[6] result_[1] separate_[1] shock_[1] southern_[1] spell_[5]
stress_[1] tense_[1] theatre_[2] therefore_[1] title_[2] tool_[1] trick_[1]
truck_[1] vary_[1]

BNC-COCA-3,000 Families: [fams 18 : types 22 : tokens 28]

analyse_[3] aspect_[1] catalogue_[2] category_[1] colony_[2] consistent_[2]
distinct_[1] elevate_[1] emphasis_[3] evident_[1] former_[1] imperial_[1]
pave_[1] rebel_[4] source_[1] sum_[1] variety_[1] versus_[1]

BNC-COCA-4,000 Families: [fams 5 : types 5 : tokens 10]

dictionary_[4] intelligent_[1] pencil_[1] plural_[1] rubber_[3]

BNC-COCA-5,000 Families: [fams 8 : types 9 : tokens 18]

condom_[1] erase_[3] flatter_[1] pillar_[1] singular_[2] usage_[2] verb_[5]
vocabulary_[3]

BNC-COCA-6,000 Families: [fams 6 : types 6 : tokens 11]

closet_[1] invert_[2] isle_[1] punctuate_[3] queue_[1] syllable_[3]

BNC-COCA-7,000 Families: [fams 1 : types 2 : tokens 3]

vowel_[3]

BNC-COCA-8,000 Families: [fams 6 : types 6 : tokens 6]

consonant_[1] jay_[1] lorry_[1] monologue_[1] pronoun_[1] slang_[1]

BNC-COCA-9,000 Families: [fams 2 : types 2 : tokens 3]

comma_[2] profane_[1]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

gregarious_[1]

BNC-COCA-11,000 Families: [fams 1 : types 2 : tokens 2]

conjugate_[2]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 1]

lolly_[1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

popside_[1]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

monolog_[1] sidewalk_[1]

7.2.2.3 *An epidemic is threatening*

7.2.2.3.1 Text only file

An epidemic is threatening

indigenous languages Marleen Haboud ,

a specialist in Andean languages , talks to Lucía Iglesias Kuntz (UNESCO) .

Community of Quichua speakers in Cotopaxi (Ecuador) .

What is the status of Central Andean languages , in terms of their viability ?

In the Central Andes (Ecuador , Peru , and Bolivia) the estimate is that one hundred indigenous languages are still alive . Determining exactly just how alive they are is not easy . This varies not only from one language to another , but also within a given language , depending on where it is spoken , the age of the speaker , their vocation , gender , level of education , etc.

For example , in Ecuador , Quechua is widely spoken in certain regions of the country , while it is rapidly disappearing from others . In this heterogeneous context , and even if certain languages continue to be used by the new generations , the general trend for all languages in the region is constant regression .

How do you explain this situation ?

Several factors are involved , such as the living conditions of native speakers , whether or not they receive institutional and social aid , the extent to which the language continues to function in all modern communication contexts , and indeed , the interest and pride of the people who speak it .

In terms of viability , the number of native speakers can be a relative concept . Some languages are spoken by a small number of people but are very much alive , such as A'i cofán in Ecuadorian Amazonia . And , on the contrary , the number of speakers of some transnational languages , such as Quechua , is dwindling every day .

Some indigenous languages maintain their vitality because of the isolation of native speakers , who find they have around them all they need to live comfortably . But isolation should not be a condition for the survival of one of these languages ; the ideal situation would be that they

cohabit with the predominant languages and societies and that they gain in strength , despite the homogenizing trends of globalization .

Why do languages disappear ? Over the last decades , a complex set of circumstances has accelerated the disappearance of indigenous languages , including contacts with other peoples , the death of native speakers , radical changes in their way of life , loss of land , massive migrations , and so on .

Only joint actions integrated with global society can curb this kind of epidemic , which is making indigenous languages and their speakers vulnerable . This presupposes that , first of all , society as a whole gets to know these languages and their speakers , and learns to respect and help keep them alive , so that we achieve the ideal of a truly multicultural society .

Another very important factor for keeping a language alive is the image that both its speakers and non speakers have of it . A person who is proud of his or her language will be more likely to keep it going .

Could you give examples of some national or regional initiatives that have helped to revitalize languages in the region ?

There have been several initiatives in our countries to help maintain minority languages . On one hand there have been government initiatives . In Andean countries , constitutional reforms have given indigenous languages an official status . The linguistic and educational policies of these countries are quite well defined and , even if they are still not always widely applied , their aim is to preserve the languages , culture and identity of their speakers , as well as respect and equality between peoples .

At the same time there are the efforts being made by speakers themselves , both collectively and individually . For example , thanks to the creation of specific family and community based educational programmes , families are trying to regain or consolidate their languages . Indigenous movements in Latin America have turned a corner in their campaign for the rights of indigenous peoples , with the creation of new bilingual , intercultural educational programmes at all levels of formal education , specific health programmes and the creation of official services

for speakers of certain languages . In some countries more than others , the media have also taken initiatives to encourage the public use of certain languages , especially those with the greatest number of speakers . Bolivia is a prime example of this .

Throughout history , new languages have been born while others have died out , why should we be concerned about the disappearance of languages ?

Just like humans themselves , languages are born and die , but we have never before seen them die at such a rapid rate as during the past decades . This means not just the loss of words and expressions , but also a store of knowledge and ways of understanding the world and communicating with others , of constructing history , of exchanging with other human beings , with elders and younger generations , and of conceptualising time , space , the living world , life and death . Each language is a universe . And , every time a word dies , unique and irreplaceable stories disappear with it .

Marleen Haboud from Ecuador is an Andean language specialist .

I speak my favourite language because that is who I am .

We teach our children our favourite language , because we want them to know who they are .

(Christine Johnson , Tohono O'odham elder , American Indian Language Development Institute , June 2002) . For a language to survive it must be passed on to the next generation .

7.2.2.3.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

That's,

3. Hyphenated words with hyphen removed:

non-speakers, community-based,

4. Compound words separated:

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Marleen, Haboud, Andean, Lucía, Iglesias, Kuntz, UNESCO, Quichua, Cotopaxi, Ecuador, Peru, Bolivia, Andes, Quechua, A'i cofán, Ecuadorian, Amazonia, America, Christine, Johnson , Tohono, O'odham, American, Indian

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Central (Andean), Central (Andes), Latin (America), (American Indian) Language Development Institute

Note: Text related to illustrations have been included in the text analysis.

7.2.2.3.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Unesco an epidemic (5.50 kb)

Words recategorized by user as 1k items (proper nouns etc): MARLEEN, HABOUD, ANDEAN, LUCÍA, IGLESIAS, KUNTZ, UNESCO, QUICHUA, COTOPAXI, ECUADOR, PERU, BOLIVIA, ANDES, QUECHUA, A'I COFÁN, ECUADORIAN, AMAZONIA, AMERICA, CHRISTINE, JOHNSON , TOHONO, O'ODHAM, AMERICAN, LATIN, INDIAN (total 56 tokens)

Families Types Tokens Percent

K1 Words (1-1000):	191	235	690	79.49%
Function:	(403)	(46.43%)
Content:	(287)	(33.06%)
> Anglo-Sax	(131)	(15.09%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	23	25	33	3.80%
> Anglo-Sax:	(12)	(1.38%)
1k+2k	(83.29%)
AWL Words (academic):	47	58	71	8.18%
> Anglo-Sax:	(3)	(0.35%)
Off-List Words:	?	53	74	8.53%
	261+?	371	868	100%

Words in text (tokens):	868
Different words (types):	371
Type-token ratio:	0.43
Tokens per type:	2.34
Lex density (content words/total)	0.54

Pertaining to onlist only

Tokens:	794
Types:	318
Families:	261
Tokens per family:	3.04
Types per family:	1.22
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
79.49	79.49
3.80	83.29
8.18	91.47
8.53	100.00

A. AWL Tokens list

AWL [47:58:71] achieve aid circumstances communicating communication community community complex concept conceptualising constant constitutional constructing contacts context contexts contrary creation creation creation culture decades decades defined despite estimate factor factors function gender generation generations generations global globalization identity image individually initiatives initiatives initiatives initiatives institute institutional integrated involved isolation isolation maintain maintain media migrations minority policies predominant prime radical region region regional regions specific specific status status survival survive trend trends unique varies

Sublist 1

concept conceptualising constitutional context contexts creation creation creation defined estimate factor factors function identity individually involved policies specific specific varies

Sublist 2

achieve community community complex constructing culture institute institutional maintain maintain region region regional regions

Sublist 3

circumstances constant minority

Sublist 4

communicating communication despite integrated status status

Sublist 5

contacts generation generations generations image prime trend trends

Sublist 6

gender initiatives initiatives initiatives initiatives migrations

Sublist 7

aid contrary decades decades global globalization isolation isolation media survival survive unique

Sublist 8

predominant radical

B. AWL Types list

AWL types: [47:58:71] achieve_[1] aid_[1] circumstances_[1] communicating_[1] communication_[1] community_[2] complex_[1] concept_[1] conceptualising_[1] constant_[1] constitutional_[1] constructing_[1] contacts_[1] context_[1] contexts_[1] contrary_[1] creation_[3] culture_[1] decades_[2] defined_[1] despite_[1] estimate_[1]

factor_1] factors_1] function_1] gender_1] generation_1] generations_2]
 global_1] globalization_1] identity_1] image_1] individually_1]
 initiatives_4] institute_1] institutional_1] integrated_1] involved_1]
 isolation_2] maintain_2] media_1] migrations_1] minority_1] policies_1]
 predominant_1] prime_1] radical_1] region_2] regional_1] regions_1]
 specific_2] status_2] survival_1] survive_1] trend_1] trends_1]
 unique_1] varies_1]

C. AWL Families list

achieve_1] aid_1] circumstance_1] communicate_2] community_2]
 complex_1] concept_2] constant_1] constitute_1] construct_1]
 contact_1] context_2] contrary_1] create_3] culture_1] decade_2]
 define_1] despite_1] estimate_1] factor_2] function_1] gender_1]
 generation_3] globe_2] identify_1] image_1] individual_1] initiate_4]
 institute_2] integrate_1] involve_1] isolate_2] maintain_2] media_1]
 migrate_1] minor_1] policy_1] predominant_1] prime_1] radical_1]
 region_4] specific_2] status_2] survive_2] trend_2] unique_1] vary_1]

AWL Fr non-cognate families: [families 2 : tokens 3] involve_1] trend_2]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, propers, cognates, extraction, barchart)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): marleen, haboud, andean, lucía, iglesias, kuntz, unesco, quichua, cotopaxi, ecuador, peru, bolivia, andes, quechua, a'ï cofán, ecuadorian, amazonia, america, christine, johnson , tohono, o'odham, american,latin indian
 end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	197 (63.55)	235 (62.83)	660 (75.77)	75.77
K-2 Words :	53 (17.10)	66 (17.65)	123 (14.12)	89.89
K-3 Words :	39 (12.58)	45 (12.03)	52 (5.97)	95.86
K-4 Words :	9 (2.90)	9 (2.41)	17 (1.95)	97.81
K-5 Words :	2 (0.65)	2 (0.53)	3 (0.34)	98.15

K-6 Words :	4 (1.29)	4 (1.07)	4 (0.46)	98.61
K-7 Words :	2 (0.65)	2 (0.53)	2 (0.23)	98.84
K-8 Words :				
K-9 Words :	1 (0.32)	1 (0.27)	1 (0.11)	98.95
K-10 Words :	1 (0.32)	1 (0.27)	1 (0.11)	99.06
K-11 Words :	1 (0.32)	1 (0.27)	1 (0.11)	99.17
K-12 Words :	1 (0.32)	1 (0.27)	1 (0.11)	99.28
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	1 (0.27)	1 (0.11)	99.39
Total (unrounded)	310+?	374 (100)	871 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	871
Different words (types):	374
Type-token ratio:	0.43
Tokens per type:	2.33

Pertaining to onlist only

Tokens:	870
Types:	373

Families: 310
 Tokens per Family : 2.81
 Types per Family : 1.20

Current profile (token %)	
K-1 (75.77)	75.77
K-2 (14.12)	89.89
K-3 (5.97)	95.86
K-4 (1.95)	97.81
K-5 (0.34)	98.15
K-6 (0.46)	98.61
K-7 (0.23)	98.84
K-9 (0.11)	98.95
K-10 (0.11)	99.06
K-11 (0.11)	99.17
K-12 (0.11)	99.28
OFF (0.11)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 165 : types 193 : tokens 630]

a [18] about [1] actions [1] age [1] all [5] also [3] always [1] am [1]
 an [3] and [30] another [2] are [11] around [1] as [7] at [3] based [1]
 be [7] because [3] been [3] before [1] being [1] beings [1] between [1]
 born [2] both [2] but [5] by [3] can [2] central [2] certain [4] changes [1]
 children [1] collectively [1] comfortably [1] concerned [1] continue [1]
 continues [1] corner [1] could [1] countries [4] country [1] day [1]
 death [2] depending [1] die [2] died [1] dies [1] do [2] during [1] each [1]
 easy [1] education [2] educational [3] especially [1] even [2] every [2]
 exactly [1] explain [1] expressions [1] families [1] family [1] favourite [2]
 find [1] first [1] for [8] from [3] general [1] gets [1] give [1] given [2]
 going [1] government [1] greatest [1] hand [1] has [1] have [11]
 health [1] help [2] helped [1] her [1] his [1] history [2] how [2] human [1]
 humans [1] hundred [1] i [3] if [2] important [1] in [19] indeed [1]
 indian [1] interest [1] involved [1] is [17] it [7] its [1] johnson [1] just [3]
 keep [2] keeping [1] kind [1] know [2] land [1] last [1] learns [1] level [1]
 levels [1] life [2] like [1] live [1] living [2] made [1] making [1] means [1]
 more [2] movements [1] much [1] must [1] my [1] national [1] need [1]
 never [1] new [3] next [1] not [6] number [5] of [47] on [5] one [4]
 only [2] or [4] other [2] others [4] our [3] out [1] over [1] passed [1]
 past [1] people [2] peoples [3] person [1] programmes [3] public [1]

quite [1] rate [1] rights [1] same [1] seen [1] services [1] set [1]
 several [2] should [2] situation [2] small [1] so [2] some [5] space [1]
 speak [2] speaker [1] speakers [14] specialist [2] spoken [3] still [2]
 store [1] stories [1] such [4] taken [1] talks [1] teach [1] terms [2]
 than [1] thanks [1] that [8] the [45] their [10] them [4] themselves [2]
 there [3] these [3] they [8] this [7] those [1] throughout [1] time [3]
 to [18] truly [1] trying [1] turned [1] understanding [1] use [1] used [1]
 very [2] want [1] way [1] ways [1] we [5] well [2] what [1] where [1]
 whether [1] which [2] while [2] who [5] whole [1] why [2] widely [2]
 will [1] with [9] within [1] word [1] words [1] world [2] would [1] you [2]
 younger [1]

BNC-COCA-2,000 types: [fams 53 : types 59 : tokens 123]

aid [1] alive [5] applied [1] circumstances [1] community [2] condition [1]
 conditions [1] constant [1] contacts [1] creation [3] culture [1]
 determining [1] development [1] disappear [2] disappearance [2]
 disappearing [1] efforts [1] elder [1] elders [1] encourage [1] equality [1]
 example [3] examples [1] exchanging [1] gain [1] generation [1]
 generations [2] identity [1] image [1] including [1] individually [1]
 intercultural [1] irreplaceable [1] june [1] knowledge [1] language [11]
 languages [25] likely [1] loss [2] maintain [2] massive [1] minority [1]
 modern [1] native [4] non [1] official [2] policies [1] pride [1] prime [1]
 proud [1] rapid [1] rapidly [1] receive [1] region [2] regional [1]
 regions [1] respect [2] social [1] societies [1] society [3] specific [2]
 strength [1] survival [1] survive [1] threatening [1] varies [1]

BNC-COCA-3,000 types: [fams 39 : types 43 : tokens 52]

accelerated [1] achieve [1] aim [1] campaign [1] communicating [1]
 communication [1] complex [1] concept [1] conceptualising [1]
 constitutional [1] constructing [1] context [1] contexts [1] decades [2]
 defined [1] despite [1] estimate [1] etc [1] extent [1] factor [1] factors [1]
 formal [1] function [1] gender [1] global [1] globalization [1] ideal [2]
 initiatives [4] institute [1] institutional [1] integrated [1] isolation [2]
 joint [1] media [1] migrations [1] preserve [1] radical [1] reforms [1]
 relative [1] status [2] trend [1] trends [1] unique [1] universe [1]
 vulnerable [1]

BNC-COCA-4,000 types: [fams 9 : types 9 : tokens 17]

consolidate [1] contrary [1] indigenous [8] linguistic [1] predominant [1]
 regain [1] regression [1] viability [2] vocation [1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 3]

curb_[1] epidemic_[2]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

dwindling_[1] multicultural_[1] presupposes_[1] vitality_[1]

BNC-COCA-7,000 types: [fams 2 : types 2 : tokens 2]

bilingual_[1] transnational_[1]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

cohabit_[1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

heterogeneous_[1]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

revitalize_[1]

BNC-COCA-12,000 types: [fams 1 : types 1 : tokens 1]

homogenizing_[1]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

B. Families list

BNC-COCA-1,000 Families: [fams 165 : types 193 : tokens 630]

a_[21] about_[1] act_[1] age_[1] all_[5] also_[3] always_[1] and_[30]
another_[2] around_[1] as_[7] at_[3] base_[1] be_[41] because_[3]
before_[1] between_[1] born_[2] both_[2] but_[5] by_[3] can_[2] centre_[2]
certain_[4] change_[1] child_[1] collect_[1] comfort_[1] concern_[1]
continue_[2] corner_[1] could_[1] country_[5] day_[1] death_[2] depend_[1]
die_[4] do_[2] during_[1] each_[1] easy_[1] educate_[5] end_of_list_[1]
especially_[1] even_[2] every_[2] exact_[1] explain_[1] express_[1]
family_[2] favourite_[2] find_[1] first_[1] for_[8] from_[3] general_[1] get_[1]
give_[3] go_[1] govern_[1] great_[1] hand_[1] have_[12] he_[1] health_[1]
help_[3] history_[2] how_[2] human_[2] hundred_[1] i_[4] if_[2] important_[1]
in_[19] indeed_[1] interest_[1] involve_[1] it_[8] just_[3] keep_[3] kind_[1]
know_[2] land_[1] last_[1] learn_[1] level_[2] life_[2] like_[1] live_[3]
make_[2] mean_[1] more_[2] move_[1] much_[1] must_[1] nation_[1]
need_[1] never_[1] new_[3] next_[1] not_[6] number_[5] of_[47] on_[5]
one_[4] only_[2] or_[4] other_[6] out_[1] over_[1] pass_[1] past_[1]
people_[5] person_[1] programme_[3] public_[1] quite_[1] rate_[1] rights_[1]
same_[1] see_[1] service_[1] set_[1] several_[2] she_[1] should_[2]
situation_[2] small_[1] so_[2] some_[5] space_[1] speak_[20] special_[2]
still_[2] store_[1] story_[1] such_[4] take_[1] talk_[1] teach_[1] term_[2]
than_[1] thank_[1] that_[9] the_[45] there_[3] they_[24] this_[10] through_[1]
time_[3] to_[18] true_[1] try_[1] turn_[1] understand_[1] use_[2] very_[2]
want_[1] way_[2] we_[8] well_[2] what_[1] where_[1] whether_[1] which_[2]
while_[2] who_[5] whole_[1] why_[2] wide_[2] will_[1] with_[9] within_[1]
word_[2] world_[2] would_[1] you_[2] young_[1]

BNC-COCA-2,000 Families: [fams 53 : types 59 : tokens 123]

aid_[1] alive_[5] apply_[1] circumstance_[1] community_[2] condition_[2]
constant_[1] contact_[1] create_[3] culture_[2] determine_[1] develop_[1]
disappear_[5] effort_[1] elder_[2] encourage_[1] equal_[1] example_[4]

exchange_[1] gain_[1] generation_[3] identify_[1] image_[1] include_[1]
individual_[1] june_[1] knowledge_[1] language_[36] likely_[1] loss_[2]
maintain_[2] massive_[1] minor_[1] modern_[1] native_[4] non_[1]
official_[2] policy_[1] pride_[1] prime_[1] proud_[1] rapid_[2] receive_[1]
region_[4] replace_[1] respect_[2] social_[1] society_[4] specific_[2]
strength_[1] survive_[2] threat_[1] vary_[1]

BNC-COCA-3,000 Families: [fams 39 : types 43 : tokens 52]

accelerate_[1] achieve_[1] aim_[1] campaign_[1] communicate_[2]
complex_[1] concept_[2] constitution_[1] construct_[1] context_[2]
decade_[2] define_[1] despite_[1] estimate_[1] etc_[1] extent_[1] factor_[2]
formal_[1] function_[1] gender_[1] global_[2] ideal_[2] initiate_[4]
institute_[1] institution_[1] integrate_[1] isolate_[2] joint_[1] media_[1]
migrate_[1] preserve_[1] radical_[1] reform_[1] relative_[1] status_[2]
trend_[2] unique_[1] universe_[1] vulnerable_[1]

BNC-COCA-4,000 Families: [fams 9 : types 9 : tokens 17]

consolidate_[1] contrary_[1] indigenous_[8] linguistic_[1] predominant_[1]
regain_[1] regress_[1] viable_[2] vocation_[1]

BNC-COCA-5,000 Families: [fams 2 : types 2 : tokens 3]

curb_[1] epidemic_[2]

BNC-COCA-6,000 Families: [fams 4 : types 4 : tokens 4]

dwindle_[1] multicultural_[1] presuppose_[1] vitality_[1]

BNC-COCA-7,000 Families: [fams 2 : types 2 : tokens 2]

bilingual_[1] transnational_[1]

BNC-COCA-8,000 Families: [fams : types : tokens]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 1]

cohabit_[1]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

heterogeneous_[1]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 1]

revitalise_[1]

BNC-COCA-12,000 Families: [fams 1 : types 1 : tokens 1]

homogenize_[1]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.2.2.4 Native Americans

7.2.2.4.1 Text only file

Starting Point

Man did not weave the web of life he is merely a strand in it Whatever he does to the web he does to himself Chief Seattle Can you think of an example where this is true

Native Americans

The indigenous people of America are known as Native Americans . They were first named " Indians " by Christopher Columbus when he arrived in America in 1492 , believing he had

reached India . Columbus not only gave indigenous Americans this name , he also helped to establish the way in which they came to be treated . Although he found them handsome and good , he was one of the first to claim their land , conquer them and take them back to the Old World as slaves .

In the USA , relations between Native Americans and Europeans have been both good and bad .

Many early colonists were helped and , in some cases , saved from certain starvation by Native Americans . Other colonists and settler were not . Although there are plenty of examples of both sides behaving well and badly , it is the Native Americans who have suffered most in this tumultuous relationship .

As mass immigration from Europe developed , relations between Europeans and Native Americans became more and more strained . The first policy of the colonists had been , basically , to leave the Indians alone . Indian land was everything east of the Appalachian Mountains and could not be bought by settlers . After the French Indian war, Indian land was defined as being the area west of the Mississippi . In 1830 Congress passed " The Indian Removal Act " ; this resulted in a mass removal of Native Americans , whose valuable land now fell into the hands of whites . In the Deep South , thousands of Native Americans belonging to the so-called civilized tribes were forced to leave their homes and settle in the " Indian Territory " , the area that is now known as Oklahoma .

These tribes were the Cherokee , Chickasaw , Choctaw , Creek and Seminole . Many perished from disease and hunger as they made their way to their new homes . One such removal , the removal of the Cherokee tribe , was later named " The Trail of Tears " . This removal was led by the US army and 4,000 of the 15,000 Cherokees died .

DO SOME MATHS

What was the death rate in this removal ? Estimate the percentage .

This ethnic removal of native populations did not only happen in the South ; it also happened in the West when the government wanted to cultivate the prairies by giving land to new immigrants .

The idea of owning land was foreign to many Native American tribes . In their view nature was sacred and belonged to everyone ; no one could buy the sky , the water , the wind or the land .

Native Americans were forced to sell the prairies they hunted on and as a result, the buffalo were shot for game and sport .

The whites' destruction of the buffalo herds meant that the Indians , who had used this animal for everything from food to tents , lost their most valuable resource .

CHECK POINT

What makes someone civilized ?

Is keeping your word a sign of civilization ?

Losing their lands disadvantaged the Native Americans in many ways . Their traditional way of living was changed to a life on reservations where the " white " ways of life were forced on them .

Today only one third of the 4.5 million Native American population lives on the reservations . In recent years these communities have increased their incomes ; primarily through gambling , but also through other types of business .

Many people of Native American descent say that they feel their Indian heritage in the way they think and feel about the world . Native American religions view the world as holy and inhabited by ancestors ; environmental issues are of great importance . These beliefs , just as the belief that one cannot buy and own land , still cause problems with the rest of society . For example , the Native American world view will consider that preserving a holy mountain is far more important than building a new ski slope . Such a perspective can be difficult for white Americans to understand .

There is a legend about the flower know as the Cherokee Rose . It sprung up whenever a tear fell to the ground ; this was to comfort the grieving tribes people on their march along the Trail of Tears . The white colour in the middle represents the gold taken from the Cherokee .

Fact file : Native Americans today

Native Americans make up 1.5 % of the total population - 4.5 million . By 2050 the population of Native Americans is projected to be 2 % of the total population of the US - 8.6 million . Native Americans are the largest ethnic minority in 5 states : Alaska , Oklahoma and South Dakota . 18 % of Alaska's total population identify themselves as Native American . 27 % speak another language than English at home . 76 % have a high school diploma . 13 % have a bachelor's degree or a higher education . 25.3 % of Native Americans are reported as living below the poverty line . 32.1 % have no life insurance . source US census bureau 2007

A guy went to a psychiatrist . Doc he said I keep having these alternating recurring dreams First I am a teepee then I am a wigwam then I am a teepee and then I am a wigwam it is driving me crazy what is wrong with me The doctor replied it is very simple you are two tents

7.2.2.4.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

I'm (4), it's (2), you're

3. Hyphenated words with hyphen removed:

French-Indian, death-rate, no-one,

4. Compound words separated:

Tribespeople, checkpoint, worldview

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Americans, America, Indian, Christopher, Columbus, India, USA, Europeans, Indians, Appalachian, French, Mississippi, Oklahoma, Cherokee, Chickasaw, Choctaw, Creek, Seminole, Cherokees, Alaska, South Dakota, Alaska's, Seattle, English

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Note: Text related to illustrations have been included in the text analysis.

7.2.2.4.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Native Americans Stunt (5.20 kb)

Words recategorized by user as 1k items (proper nouns etc): AMERICANS, AMERICA, INDIAN, CHRISTOPHER, COLUMBUS, INDIA, USA, EUROPEANS, INDIANS, APPALACHIAN, FRENCH, MISSISSIPPI, OKLAHOMA, CHEROKEE, CHICKASAW, CHOCTAW, CREEK, SEMINOLE, CHEROKEES, ALASKA, SOUTH DAKOTA, ALASKA'S, SEATTLE, ENGLISH, (total 57 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	218	274	734	81.56%
Function:	(406)	(45.11%)
Content:	(328)	(36.44%)
> Anglo-Sax	(172)	(19.11%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	23	24	30	3.33%
> Anglo-Sax:	(11)	(1.22%)
1k+2k		(84.89%)
AWL Words (academic):	23	24	32	3.56%
> Anglo-Sax:	(7)	(0.78%)
Off-List Words:	<u>2</u>	<u>60</u>	<u>104</u>	<u>11.56%</u>
	264+?	382	900	100%

Words in text (tokens):	900
Different words (types):	382
Type-token ratio:	0.42
Tokens per type:	2.36
Lex density (content words/total)	0.55

Pertaining to onlist only

Tokens:	796
Types:	322

Families:	264
Tokens per family:	3.02
Types per family:	1.22
Anglo-Sax Index: <small>(A-Sax tokens + functors / onlist tokens)</small>	%
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
81.56	81.56
3.33	84.89
3.56	88.45
11.56	100.00

A. AWL Tokens lists

AWL [23:24:32] alternating area area communities defined environmental establish estimate ethnic ethnic file identify immigrants immigration incomes issues minority percentage perspective policy primarily projected removal removal removal removal removal removal resource source traditional

Sublist 1

area area defined environmental establish estimate identify incomes issues percentage policy source

Sublist 2

communities primarily resource traditional

Sublist 3

immigrants immigration minority removal removal removal removal removal removal removal removal

Sublist 4

ethnic ethnic projected

Sublist 5

alternating perspective

Sublist 7

file

B. AWL Types list

AWL types: [23:24:32] alternating_[1] area_[2] communities_[1] defined_[1] environmental_[1] establish_[1] estimate_[1] ethnic_[2] file_[1] identify_[1] immigrants_[1] immigration_[1] incomes_[1] issues_[1] minority_[1] percentage_[1] perspective_[1] policy_[1] primarily_[1] projected_[1] removal_[7] resource_[1] source_[1] traditional_[1]

C. AWL Families list

AWL families: [23:24:32]

alter_[1] area_[2] community_[1] define_[1] environment_[1] establish_[1] estimate_[1] ethnic_[2] file_[1] identify_[1] immigrate_[2] income_[1] issue_[1] minor_[1] percent_[1] perspective_[1] policy_[1] primary_[1] project_[1] remove_[7] resource_[1] source_[1] tradition_[1]

AWL Fr non-cognate families: [families 1 : tokens 7] remove_[7]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, propers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: Native Americans Stunt (5,390 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): americans, america, indian, christopher, columbus, india, usa, europeans, indians, appalachian, french, mississippi, oklahoma, cherokee, chickasaw, choctaw, creek, seminole, cherokees, alaska, south dakota, alaska's, seattle, english, end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	228 (71.03)	277 (72.32)	739 (82.11)	82.11
K-2 Words :	38 (11.84)	42 (10.97)	78 (8.67)	90.78
K-3 Words :	31 (9.66)	34 (8.88)	43 (4.78)	95.56
K-4 Words :	13 (4.05)	13 (3.39)	14 (1.56)	97.12

K-5 Words :	2 (0.62)	2 (0.52)	2 (0.22)	97.34
K-6 Words :	4 (1.25)	4 (1.04)	4 (0.44)	97.78
K-7 Words :	2 (0.62)	2 (0.52)	4 (0.44)	98.22
K-8 Words :	1 (0.31)	1 (0.26)	1 (0.11)	98.33
K-9 Words :				
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	1 (0.31)	1 (0.26)	2 (0.22)	98.55
K-16 Words :				
K-17 Words :	1 (0.31)	1 (0.26)	2 (0.22)	98.77
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	0 (0.00)	0 (0.00)	
Total (unrounded)	321+?	383 (100)	900 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	900
Different words (types):	383
Type-token ratio:	0.43
Tokens per type:	2.35

Pertaining to onlist only

Tokens:	900
---------	-----

Types: 383
 Families: 321
 Tokens per Family : 2.80
 Types per Family : 1.19

Current profile (token %)		
K-1	(82.11)	82.11
K-2	(8.67)	90.78
K-3	(4.78)	95.56
K-4	(1.56)	97.12
K-5	(0.22)	97.34
K-6	(0.44)	97.78
K-7	(0.44)	98.22
K-8	(0.11)	98.33
K-15	(0.22)	98.55
K-17	(0.22)	98.77
OFF	(0.00)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 192 : types 227 : tokens 689]

a_[19] about_[2] act_[1] after_[1] alone_[1] along_[1] also_[3] although_[2]
 am_[4] an_[1] and_[22] animal_[1] another_[1] are_[6] area_[2] arrived_[1]
 as_[12] at_[1] back_[1] bad_[1] badly_[1] basically_[1] be_[4] became_[1]
 been_[2] being_[1] believing_[1] below_[1] between_[2] both_[2] bought_[1]
 building_[1] business_[1] but_[1] buy_[2] by_[7] called_[1] came_[1] can_[2]
 cannot_[1] cases_[1] cause_[1] certain_[1] changed_[1] check_[1]
 colour_[1] comfort_[1] consider_[1] could_[2] crazy_[1] death_[1] deep_[1]
 degree_[1] did_[2] died_[1] difficult_[1] do_[1] doc_[1] doctor_[1] does_[2]
 dreams_[1] driving_[1] early_[1] east_[1] education_[1] everyone_[1]
 everything_[2] fact_[1] far_[1] feel_[2] fell_[2] first_[4] flower_[1] food_[1]
 for_[4] forced_[3] found_[1] from_[5] game_[1] gave_[1] giving_[1] gold_[1]
 good_[2] government_[1] great_[1] ground_[1] guy_[1] had_[3] hands_[1]
 happen_[1] happened_[1] have_[6] having_[1] he_[9] helped_[2] high_[1]
 higher_[1] himself_[1] home_[1] homes_[2] hunger_[1] hunted_[1] i_[5]
 idea_[1] important_[1] in_[20] insurance_[1] into_[1] is_[11] issues_[1] it_[6]
 just_[1] keep_[1] keeping_[1] know_[1] known_[2] land_[8] lands_[1]
 largest_[1] later_[1] leave_[2] led_[1] life_[4] line_[1] lives_[1] living_[2]
 losing_[1] lost_[1] made_[2] make_[1] makes_[1] man_[1] many_[5] me_[2]
 meant_[1] middle_[1] million_[3] more_[3] most_[2] mountain_[1]

mountains_[1] name_[1] named_[2] nature_[1] new_[3] no_[2] not_[5]
 now_[2] number_[26] of_[33] old_[1] on_[5] one_[5] only_[3] or_[2] other_[2]
 own_[1] owning_[1] passed_[1] people_[3] plenty_[1] point_[2] problems_[1]
 rate_[1] reached_[1] recent_[1] relations_[2] relationship_[1] replied_[1]
 reported_[1] rest_[1] rose_[1] said_[1] saved_[1] say_[1] school_[1] sell_[1]
 settle_[1] settler_[1] settlers_[1] shot_[1] sides_[1] sign_[1] simple_[1]
 sky_[1] so_[1] some_[2] someone_[1] south_[3] speak_[1] sport_[1]
 sprung_[1] starting_[1] still_[1] such_[2] take_[1] taken_[1] tear_[1] tears_[2]
 than_[2] that_[5] the_[69] their_[11] them_[5] themselves_[1] then_[2]
 there_[2] these_[4] they_[6] think_[2] third_[1] this_[9] thousands_[1]
 through_[2] to_[22] today_[2] total_[3] treated_[1] true_[1] two_[1] types_[1]
 understand_[1] up_[2] us_[3] used_[1] very_[1] view_[3] wanted_[1] war_[1]
 was_[10] water_[1] way_[4] ways_[2] web_[2] well_[1] went_[1] were_[8]
 west_[2] what_[3] whatever_[1] when_[2] whenever_[1] where_[2] which_[1]
 white_[4] whites_[1] who_[2] whose_[1] will_[1] wind_[1] with_[2] word_[1]
 world_[4] wrong_[1] years_[1] you_[2] your_[1]

BNC-COCA-2,000 types: [fams 38 : types 40 : tokens 78]

army_[1] belonged_[1] belonging_[1] chief_[1] claim_[1] communities_[1]
 developed_[1] disease_[1] environmental_[1] establish_[1] example_[2]
 examples_[1] file_[1] foreign_[1] identify_[1] incomes_[1] increased_[1]
 language_[1] march_[1] mass_[2] maths_[1] minority_[1] native_[22]
 percentage_[1] policy_[1] population_[5] populations_[1] projected_[1]
 removal_[7] represents_[1] reservations_[2] result_[1] resulted_[1] ski_[1]
 slaves_[1] society_[1] starvation_[1] states_[1] suffered_[1] tents_[2]
 traditional_[1] valuable_[2]

BNC-COCA-3,000 types: [fams 30 : types 31 : tokens 43]

behaving_[1] belief_[1] beliefs_[1] bureau_[1] civilization_[1] civilized_[2]
 colonists_[3] congress_[1] defined_[1] destruction_[1] estimate_[1]
 ethnic_[2] heritage_[1] holy_[2] immigrants_[1] importance_[1] inhabited_[1]
 legend_[1] merely_[1] perspective_[1] poverty_[1] preserving_[1]
 primarily_[1] psychiatrist_[1] religions_[1] resource_[1] slope_[1] source_[1]
 strained_[1] territory_[1] trail_[2] tribe_[1] tribes_[4] weave_[1]

BNC-COCA-4,000 types: [fams 13 : types 13 : tokens 14]

alternating_[1] ancestors_[1] census_[1] cultivate_[1] disadvantaged_[1]
 gambling_[1] handsome_[1] herds_[1] immigration_[1] indigenous_[2]
 recurring_[1] sacred_[1] strand_[1]

BNC-COCA-5,000 types: [fams 3 : types 3 : tokens 3]

conquer_[1] creek_[1] descent_[1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

bachelor_[1] diploma_[1] grieving_[1] perished_[1]

BNC-COCA-7,000 types: [fams 2 : types 2 : tokens 4]

buffalo_[2] prairies_[2]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

reciprocal_[1]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 2]

teepee_[2]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens]

wigwam_[2]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

B. Families list

BNC-COCA-1,000 Families: [fams 192 : types 227 : tokens 689]

a_[20] about_[2] act_[1] after_[1] alone_[1] along_[1] also_[3] although_[2]
and_[22] animal_[1] another_[1] area_[2] arrive_[1] as_[12] at_[1] back_[1]
bad_[2] basic_[1] be_[46] become_[1] believe_[1] below_[1] between_[2]
both_[2] build_[1] business_[1] but_[1] buy_[3] by_[7] call_[1] can_[3]
case_[1] cause_[1] certain_[1] change_[1] check_[1] colour_[1] come_[1]
comfort_[1] consider_[1] could_[2] crazy_[1] death_[1] deep_[1] degree_[1]
die_[1] difficult_[1] do_[5] doctor_[2] dream_[1] drive_[1] early_[1] east_[1]
educate_[1] every_[3] fact_[1] fall_[2] far_[1] feel_[2] find_[1] first_[4]
flower_[1] food_[1] for_[4] force_[3] from_[5] game_[1] give_[2] go_[1]
gold_[1] good_[2] govern_[1] great_[1] ground_[1] guy_[1] hand_[1]
happen_[2] have_[10] he_[10] help_[2] high_[2] home_[3] hunger_[1]
hunt_[1] i_[7] idea_[1] important_[1] in_[20] insure_[1] into_[1] issue_[1]
it_[6] just_[1] keep_[2] know_[3] land_[9] large_[1] late_[1] lead_[1]
leave_[2] life_[4] line_[1] live_[3] lose_[2] make_[3] man_[1] many_[5]
mean_[1] middle_[1] million_[3] more_[3] most_[2] mountain_[2] name_[3]
nature_[1] new_[3] no_[2] not_[5] now_[2] number_[26] of_[33] old_[1]
on_[5] one_[5] only_[3] or_[2] other_[2] own_[1] owned_[1] pass_[1]
people_[3] plenty_[1] point_[2] problem_[1] rate_[1] reach_[1] recent_[1]
relate_[3] reply_[1] report_[1] rest_[1] rise_[1] save_[1] say_[2] school_[1]
sell_[1] settle_[3] shoot_[1] side_[1] sign_[1] simple_[1] sky_[1] so_[1]
some_[3] south_[3] speak_[1] sport_[1] spring_[1] start_[1] still_[1] such_[2]
take_[2] tear_[3] than_[2] that_[5] the_[69] then_[2] there_[2] they_[23]
think_[2] this_[13] thousand_[1] three_[1] through_[2] to_[22] today_[2]
total_[3] treat_[1] true_[1] two_[1] type_[1] understand_[1] up_[2] use_[1]
very_[1] view_[3] want_[1] war_[1] water_[1] way_[6] we_[3] web_[2]
well_[1] west_[2] what_[4] when_[3] where_[2] which_[1] white_[5] who_[3]
will_[1] wind_[1] with_[2] word_[1] world_[4] wrong_[1] year_[1] you_[3]

BNC-COCA-2,000 Families: [fams 38 : types 40 : tokens 78]

army_[1] belong_[2] chief_[1] claim_[1] community_[1] develop_[1]
disease_[1] environment_[1] establish_[1] example_[3] file_[1] foreign_[1]
identify_[1] income_[1] increase_[1] language_[1] march_[1] mass_[2]
mathematics_[1] minor_[1] native_[22] percent_[1] policy_[1] population_[6]

project_[1] remove_[7] represent_[1] reserve_[2] result_[2] ski_[1] slave_[1]
society_[1] starve_[1] states_[1] suffer_[1] tent_[2] tradition_[1] value_[2]

BNC-COCA-3,000 Families: [fams 30 : types 31 : tokens 43]

behave_[1] belief_[2] bureau_[1] civilise_[3] colony_[3] congress_[1]
define_[1] destruction_[1] estimate_[1] ethnic_[2] heritage_[1] holy_[2]
immigrant_[1] importance_[1] inhabit_[1] legend_[1] mere_[1]
perspective_[1] poverty_[1] preserve_[1] primary_[1] psychiatry_[1]
religion_[1] resource_[1] slope_[1] source_[1] strain_[1] territory_[1] trail_[2]
tribe_[5] weave_[1]

BNC-COCA-4,000 Families: [fams 13 : types 13 : tokens 14]

alternate_[1] ancestor_[1] census_[1] cultivate_[1] disadvantage_[1]
gamble_[1] handsome_[1] herd_[1] immigrate_[1] indigenous_[2] recur_[1]
sacred_[1] strand_[1]

BNC-COCA-5,000 Families: [fams 3 : types 3 : tokens 3]

conquer_[1] creek_[1] descent_[1]

BNC-COCA-6,000 Families: [fams 4 : types 4 : tokens 4]

bachelor_[1] diploma_[1] grieve_[1] perish_[1]

BNC-COCA-7,000 Families: [fams 2 : types 2 : tokens 4]

buffalo_[2] prairie_[2]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

tumult_[1]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 2]

teepee_[2]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 2

wigwam_[2]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.2.2.5 *Australia: The Birth of a Nation*

7.2.2.5.1 Text only file

Starting Point

What are the reasons people leave their home to move to another country ?

Sydney and Melbourne were the two dominant cities in Australia at the time of the nation's birth , in 1901 . Powerful voices in both New South Wales and Victoria argued that their capital city should be the new national capital . However , instead of choosing one of the two , Parliament decided to construct a new capital , on a site between Sydney and Melbourne . An inland area by the Monlonglo river was chosen and they named the place Canberra .

Australia : The Birth of a Nation

The Australian continent has been inhabited for more than 40,000 years , but Australian society as we know it has only been created over the last two centuries .

The Industrial Revolution brought about great changes in British society in the 18 and 19 centuries. More people made more money than before , but at the same time the gap between the rich and the poor widened . The crime rate also

" True patriots we ; for be it understood , We left our country for our country's good . "

Henry Carter died in 1806

rose ; it is estimated that in the late 1700s more than 10 % of the population of London made its living from criminal activity . For a long time criminals had been shipped across the Atlantic to the colonies in North America , but when the Treaty of Paris was signed in 1783 , making the US an independent nation , this had to be stopped .

The USA refused to function as a dumping ground for the criminal class . The British needed a new place to send their prisoners , and , after much consideration , they chose Botany Bay in New South Wales .

It is believed that the first European explorers reached Australia in 1606 and several ships visited the continent in the following years . But it was Captain James Cook , on an expedition in 1770 , who claimed possession of the east coast on behalf of the English king , George . He named the region New South Wales . The first Australian colony was established in 1788 when the first fleet of British settlers and convicts arrived .

The settlers learnt to cultivate the land , small towns were established and life , eventually , became quite comfortable for many of them . As time went by , many of the convicts who had served their sentences were also allowed to settle in the colony , as free men . There was plenty of land so they had the opportunity to start new lives . In the beginning of the 19 century , stories about large Australian farms and the country's wealth spread to England and newspapers wrote about the land of opportunities . Many people were tempted . They left England behind and sailed to Australia as free settlers . The colony's " upper class " was happy with this new wave of immigration , and the new inhabitants were given large areas of land.

In 1851 , gold was found in New South Wales and there were several other discoveries in other parts of the country . The news of the findings spread around the world and fortune hunters came from everywhere . Thus , it was no longer considered much of a punishment to be deported to the colonies down under the very last convict ship carrying Irish prisoners , arrived in Perth in 1868 In all 162,000 convicts were transported to the Australian colonies .

From the very beginning , when the first convicts were sent , Australia was very much a masculine country . The groups of convicts that were deported to this corner of the world mainly consisted of men and many of them had probably lived rough lives on the streets and in city slums . This was reflected in their behaviour . Many of this large group of bachelors , with nothing else to do in their spare time , spent their evenings drinking and gambling . In time , they were joined by the many men who came to Australia to seek their fortunes in the gold mines .

There were not enough women , and something had to be done about this . Advertisements were published in British papers , in the late 19 century , to encourage middle class women to emigrate to the new colonies . Cheap fares were offered to those who wanted to go . There was a need for house keepers , teachers and wives . Britain had a surplus of women at this time , so many single women saw this as an opportunity to find work or even better a husband .

Towards the end of the 19 century , Australia consisted of six colonies - New South Wales , Victoria , South Australia , Queensland, Western Australia and Tasmania . As the population grew and immigrants started to come from different corners of the world , the settlers of British descent were worried that they would have to give up their power . By this time , a large number of the inhabitants had been born and raised in the colonies . They had never been to Europe and did not consider themselves British in the way their parents did . Thus , the six colonies came together and after a referendum the commonwealth of Australia was established in 1901 .

Captain Cook takes New South Wales

7.2.2.5.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

middle-class

4. Compound words separated:

housekeepers

5. Words (groups of letters) removed from the text analysis:

18th, 19th (th removed), George (III)

6. Proper nouns:

Sydney, Melbourne, Australia, Wales, Victoria, Canberra, British, Henry, Carter, Cook, America, Paris, USA, British, Botany, European, James, English, Australian, George, England, Perth, Queensland, Tasmania, Europe, Britain, Irish, London, Atlantic, Monlonglo

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Note: Text related to illustrations have been included in the text analysis.

7.2.2.5.3 Text analysis

Text Analysis: Stunt - Australia - Birth of a Nation

1. VP-Classic

WEB VP OUTPUT FOR FILE: Australia - Birth of a Nation (4.92 kb)

Words recategorized by user as 1k items (proper nouns etc): SYDNEY, MELBOURNE, AUSTRALIA, WALES, VICTORIA, CANBERRA, BRITISH, HENRY, CARTER, COOK, AMERICA, PARIS, USA, BRITISH, BOTANY, EUROPEAN, JAMES, ENGLISH, AUSTRALIAN, GEORGE,

ENGLAND, PERTH, QUEENSLAND, TASMANIA, EUROPE, BRITAIN, IRISH, LONDON, ATLANTIC, MONLONGLO (total 57 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	208	255	704	83.02%
Function:	(403)	(47.52%)
Content:	(301)	(35.50%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(147)	(17.33%)
K2 Words (1001-2000):	28	30	37	4.36%
> Anglo-Sax:	(13)	(1.53%)
1k+2k	(87.38%)
AWL Words (academic):	17	19	22	2.59%
> Anglo-Sax:	(2)	(0.24%)
Off-List Words:	<u>?</u>	<u>51</u>	<u>85</u>	<u>10.02%</u>
	253+?	355	848	100%

Words in text (tokens):	848
Different words (types):	355
Type-token ratio:	0.42
Tokens per type:	2.39
Lex density (content words/total)	0.52

Pertaining to onlist only

Tokens:	763
Types:	304
Families:	253
Tokens per family:	3.02
Types per family:	1.20
Anglo-Sax Index: <small>(A-Sax tokens + functors / onlist tokens)</small>	%
Greco-Lat/Fr-Cognate Index: <small>(inverse of above)</small>	%

Current profile	
%	Cumul.
83.02	83.02
4.36	87.38
2.59	89.97
10.02	100.00

A. AWL Tokens lists

AWL [17:19:22] area areas behalf consisted consisted construct created dominant established established established estimated eventually function immigrants immigration published region revolution seek site transported

Sublist 1

area areas consisted consisted created established established established estimated function

Sublist 2

construct region seek site

Sublist 3

dominant immigrants immigration published

Sublist 6

transported

Sublist 8

eventually

Sublist 9

behalf revolution

B. AWL Types list

AWL types: [17:19:22] area_[1] areas_[1] behalf_[1] consisted_[2] construct_[1] created_[1] dominant_[1] established_[3] estimated_[1] eventually_[1] function_[1] immigrants_[1] immigration_[1] published_[1] region_[1] revolution_[1] seek_[1] site_[1] transported_[1]

C. AWL Families list

area_[2] behalf_[1] consist_[2] construct_[1] create_[1] dominate_[1] establish_[3] estimate_[1] eventual_[1] function_[1] immigrate_[2] publish_[1] region_[1] revolution_[1] seek_[1] site_[1] transport_[1]

AWL Fr non-cognate families: [families 2 : tokens 2] behalf_[1] seek_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: Australia - Birth of a Nation (5,043 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): sydney, melbourne, australia, wales, victoria, canberra, british, henry, carter, cook, america, paris, usa, british, botany, european, james, english, australian, george, england, perth, queensland, tasmania, europe, britain, irish, london, atlantic, monlonglo end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	229 (75.83)	273 (76.69)	736 (86.79)	86.79
K-2 Words :	33 (10.93)	37 (10.39)	51 (6.01)	92.80
K-3 Words :	22 (7.28)	25 (7.02)	40 (4.72)	97.52
K-4 Words :	9 (2.98)	9 (2.53)	9 (1.06)	98.58
K-5 Words :	6 (1.99)	6 (1.69)	6 (0.71)	99.29
K-6 Words :	3 (0.99)	3 (0.84)	4 (0.47)	99.76
K-7 Words :				
K-8 Words :				
K-9 Words :				
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				

K-24 Words :

K-25 Words :

Off-List: ?? 0 (0.00) 0 (0.00)

Total (unrounded) 302+? 356 (100) 848 (100) ≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens): 848

Different words (types): 356

Type-token ratio: 0.42

Tokens per type: 2.38

Pertaining to onlist only

Tokens: 848

Types: 356

Families: 302

Tokens per Family : 2.81

Types per Family : 1.18

Current profile (token %)	
K-1 (86.79)	86.79
K-2 (6.01)	92.80
K-3 (4.72)	97.52
K-4 (1.06)	98.58
K-5 (0.71)	99.29
K-6 (0.47)	99.76
OFF (0.00)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 181 : types 211 : tokens 684]

a_[13] about_[4] across_[1] advertisements_[1] after_[2] all_[1] allowed_[1] also_[2] an_[4] and_[26] another_[1] are_[1] area_[1] areas_[1] around_[1] arrived_[2] as_[7] at_[3] be_[5] became_[1] been_[5] before_[1] beginning_[2] behind_[1] believed_[1] better_[1] between_[2] birth_[2] born_[1] both_[1] brought_[1] but_[4] by_[4] came_[3] carrying_[1] changes_[1] cheap_[1] choosing_[1] chose_[1] chosen_[1] cities_[1] city_[2] class_[3] come_[1] comfortable_[1] consider_[1] consideration_[1] considered_[1] cook_[2] corner_[1] corners_[1] country_[6] crime_[1] decided_[1] did_[2] died_[1] different_[1] discoveries_[1] do_[1] done_[1] down_[1] drinking_[1] east_[1] else_[1] end_[1] enough_[1] even_[1] evenings_[1] everywhere_[1] farms_[1] find_[1] findings_[1] first_[4] following_[1] for_[7] found_[1] free_[2] from_[4] give_[1] given_[1] go_[1] gold_[2] good_[1] great_[1] grew_[1] ground_[1] group_[1] groups_[1] had_[9] happy_[1] has_[2] have_[1] he_[1] home_[1] house_[1] however_[1] hunters_[1] husband_[1] in_[32] instead_[1] is_[2] it_[6] its_[1] joined_[1] keepers_[1] king_[1] know_[1] land_[4] large_[4] last_[2] late_[2] learnt_[1] leave_[1] left_[2] life_[1] lived_[1] lives_[2] living_[1] long_[1] longer_[1] made_[2] mainly_[1] making_[1] many_[7] men_[3] middle_[1] money_[1] monlonglo_[1] more_[4] move_[1] much_[3] named_[2] nation_[3] national_[1] need_[1] needed_[1] never_[1] new_[13] news_[1] no_[1] north_[1] not_[2] nothing_[1] number_[20] numbers_[1] of_[33] offered_[1] on_[4] one_[1] only_[1] or_[1] other_[2] our_[2] over_[1] papers_[1] parents_[1] parts_[1] people_[3] place_[2] plenty_[1] point_[1] poor_[1] power_[1] powerful_[1] prisoners_[2] probably_[1] quite_[1] raised_[1] rate_[1] reached_[1] reasons_[1] rich_[1] river_[1] rose_[1] rough_[1] sailed_[1] same_[1] saw_[1] send_[1] sent_[1] served_[1] settle_[1] settlers_[4] several_[2] ship_[1] shipped_[1] ships_[1] should_[1] signed_[1] single_[1] six_[2] small_[1] so_[2] something_[1] south_[7] spent_[1] start_[1] started_[1] starting_[1] stopped_[1] stories_[1] streets_[1] takes_[1] teachers_[1] than_[3] that_[5] the_[72] their_[10] them_[2] themselves_[1] there_[4] they_[7] this_[9] those_[1] time_[8] to_[29] together_[1] towards_[1] towns_[1] true_[1] two_[3] under_[1] understood_[1] up_[1] us_[1] very_[3] visited_[1] voices_[1] wanted_[1] was_[12] wave_[1] way_[1] we_[3] went_[1] were_[14] what_[1] when_[3] who_[4] widened_[1] with_[2] wives_[1] women_[4] work_[1] world_[3] worried_[1] would_[1] wrote_[1] years_[2]

BNC-COCA-2,000 types: [fams 34 : types 38 : tokens 52]

activity_[1] argued_[1] bay_[1] capital_[3] captain_[2] carter_[1] centuries_[2] century_[3] claimed_[1] coast_[1] created_[1] criminal_[2] criminals_[1] dumping_[1] encourage_[1] established_[3] eventually_[1] fortune_[1] fortunes_[1] industrial_[1] newspapers_[1] opportunities_[1] opportunity_[2] population_[2] possession_[1] punishment_[1] refused_[1]

region_[1] seek_[1] sentences_[1] site_[1] society_[2] spare_[1] spread_[2] tempted_[1] thus_[2] upper_[1] western_[1]

BNC-COCA-3,000 types: [fams 22 : types 24 : tokens 40]

behaviour_[1] colonies_[7] colony_[3] consisted_[2] construct_[1] continent_[2] convict_[1] convicts_[5] dominant_[1] estimated_[1] explorers_[1] function_[1] gap_[1] immigrants_[1] independent_[1] inhabitants_[2] inhabited_[1] mines_[1] parliament_[1] published_[1] reflected_[1] revolution_[1] transported_[1] treaty_[1] wealth_[1]

BNC-COCA-4,000 types: [fams 9 : types 9 : tokens 9]

behalf_[1] cultivate_[1] expedition_[1] fares_[1] fleet_[1] gambling_[1] immigration_[1] patriots_[1] surplus_[1]

BNC-COCA-5,000 types: [fams 7 : types 7 : tokens 7]

botany_[1] commonwealth_[1] descent_[1] emigrate_[1] inland_[1] masculine_[1] referendum_[1]

BNC-COCA-6,000 types: [fams 3 : types 3 : tokens 4]

bachelors_[1] deported_[2] slums_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

B. Families list

BNC-COCA-1,000 Families: [fams 181 : types 211 : tokens 684]

a_[17] about_[4] across_[1] advertise_[1] after_[2] all_[1] allow_[1] also_[2]
and_[26] another_[1] area_[2] around_[1] arrive_[2] as_[7] at_[3] be_[39]
become_[1] before_[1] begin_[2] behind_[1] believe_[1] better_[1]
between_[2] birth_[2] born_[1] both_[1] bring_[1] but_[4] by_[4] carry_[1]
change_[1] cheap_[1] choose_[3] city_[3] class_[3] come_[4] comfort_[1]
consider_[3] cook_[2] corner_[2] country_[6] crime_[1] decide_[1] die_[1]
different_[1] discover_[1] do_[4] down_[1] drink_[1] east_[1] else_[1]
end_[1] end_of_list_[1] enough_[1] even_[1] evening_[1] every_[1] farm_[1]
find_[3] first_[4] follow_[1] for_[7] free_[2] from_[4] give_[2] go_[2] gold_[2]
good_[1] great_[1] ground_[1] group_[2] grow_[1] happy_[1] have_[12]
he_[1] home_[1] house_[1] however_[1] hunt_[1] husband_[1] in_[32]
instead_[1] it_[7] join_[1] keep_[1] king_[1] know_[1] land_[4] large_[4]
last_[2] late_[2] learn_[1] leave_[1] left_[2] life_[1] live_[4] long_[2] main_[1]
make_[3] man_[3] many_[7] middle_[1] money_[1] more_[4] move_[1]
much_[3] name_[2] nation_[4] need_[2] never_[1] new_[13] news_[1] no_[1]
north_[1] not_[2] nothing_[1] number_[21] of_[33] offer_[1] on_[4] one_[1]
only_[1] or_[1] other_[2] over_[1] paper_[1] parent_[1] part_[1] people_[3]
place_[2] plenty_[1] point_[1] poor_[1] power_[2] prison_[2] probably_[1]
quite_[1] raise_[1] rate_[1] reach_[1] reason_[1] rich_[1] rise_[1] river_[1]
rough_[1] sail_[1] same_[1] see_[1] send_[2] serve_[1] settle_[5]

several_[2] ship_[3] should_[1] sign_[1] single_[1] six_[2] small_[1] so_[2]
some_[1] south_[7] spend_[1] start_[3] stop_[1] story_[1] street_[1] take_[1]
teach_[1] than_[3] that_[6] the_[72] there_[4] they_[20] this_[9] time_[8]
to_[29] together_[1] toward_[1] town_[1] true_[1] two_[3] under_[1]
understand_[1] up_[1] very_[3] visit_[1] voice_[1] want_[1] wave_[1] way_[1]
we_[6] what_[1] when_[3] who_[4] wide_[1] wife_[1] with_[2] woman_[4]
work_[1] world_[3] worry_[1] would_[1] write_[1] year_[2]

BNC-COCA-2,000 Families: [fams 34 : types 38 : tokens 52]

active_[1] argue_[1] bay_[1] capital_[3] captain_[2] cart_[1] century_[5]
claim_[1] coast_[1] create_[1] criminal_[3] dump_[1] encourage_[1]
establish_[3] eventually_[1] fortune_[2] industry_[1] newspaper_[1]
opportunity_[3] population_[2] possess_[1] punish_[1] refuse_[1] region_[1]
seek_[1] sentence_[1] site_[1] society_[2] spare_[1] spread_[2] tempt_[1]
thus_[2] upper_[1] western_[1]

BNC-COCA-3,000 Families: [fams 22 : types 24 : tokens 40]

behaviour_[1] colony_[10] consist_[2] construct_[1] continent_[2] convict_[6]
dominant_[1] estimate_[1] explore_[1] function_[1] gap_[1] immigrant_[1]
independent_[1] inhabit_[3] miner_[1] parliament_[1] publish_[1] reflect_[1]
revolution_[1] transport_[1] treaty_[1] wealth_[1]

BNC-COCA-4,000 Families: [fams 9 : types 9 : tokens 9]

behalf_[1] cultivate_[1] expedition_[1] fare_[1] fleet_[1] gamble_[1]
immigrate_[1] patriot_[1] surplus_[1]

BNC-COCA-5,000 Families: [fams 7 : types 7 : tokens 7]

botany_[1] commonwealth_[1] descent_[1] emigrate_[1] inland_[1]
masculine_[1] referendum_[1]

BNC-COCA-6,000 Families: [fams 3 : types 3 : tokens 4]

bachelor_[1] deport_[2] slum_[1]

BNC-COCA-7,000 Families: [fams : types : tokens]

BNC-COCA-8,000 Families: [fams : types : tokens]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.2.2.6 *Stolen Generation*

7.2.2.6.1 Text only file

Starting Point

In some cases it is necessary to take children away from their parents to protect them and give them better opportunities . Under which circumstances do you think it is right to take a child away from its parents ?

Stolen Generation

When the first Europeans reached the shores of Australia in 1606 , the country had already been inhabited for thousands of years by the Aborigines , who had arrived 40,000 - 50,000 years earlier from the Asian main land and the islands north of the Australian continent .

From the very beginning , the Europeans were a threat to these indigenous people . The first settlers brought unknown diseases ; smallpox was

one of the reasons why the native population was , in the course of a few years , halved. The Aborigines were also driven away from the most fertile land as the whites wanted to cultivate these areas and settle there .

The European settlers regarded Aborigines as an inferior race . It was difficult to understand that these people could be happy with their simple , nomadic life style . They did not settle in one place , they were on the move all the time . This life style made it difficult to have a proper home , go to school and have a permanent job . In addition , they had their own cultural rituals and religions . They were perceived as a threat to the Europeans , who were trying to establish a new nation on the Australian continent .

The white government did not approve of the Aboriginal life style . As a consequence , the authorities started removing Aboriginal children from their homes in the early 1860s . These children were taken to orphanages or foster homes . The first official legislation , however , is to be found in the Victorian Aboriginal Protection Act of 1869 . For more than 100 years , children of Australian Aboriginal and Torres Strait Islander descent were removed from their families by the government or the church . These children are referred to as the " stolen generation " . They were taken away from their natural environment and did not get the chance to know their family and their culture . The last children were removed

from their homes in the 1970s . The official excuse for doing this was to protect the children from neglect and abuse . It is estimated that more than 100,000 children were taken away from their parents during these years and not one single Aboriginal family was spared .

An official debate on the sufferings these people experienced was not started until the mid 1990s . This debate led to demands for an official apology . Neither Prime Minister John Howard nor his predecessor Prime Minister Paul Keating agreed to make one . It was finally made when Kevin Rudd , shortly after becoming the country's prime minister , gave an official apology in February 2008 .

7.2.2.6.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

mid-1990s

4. Compound words separated:

lifestyle (3), mainland

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Europeans, Australia, Aborigines, Asian, Australian, European, Aboriginal, Victorian, Torres, Strait, Islander, John, Howard, Paul, Keating, Kevin, Rudd,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Note: Text related to illustrations have been included in the text analysis.

7.2.2.6.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Stolen Generation stunt (2.76 kb)

Words recategorized by user as 1k items (proper nouns etc): EUROPEANS, AUSTRALIA, ABORIGINES, ASIAN, AUSTRALIAN, EUROPEAN, ABORIGINAL, VICTORIAN, TORRES, STRAIT, ISLANDER, JOHN, HOWARD, PAUL, KEATING, KEVIN, RUDD, (total 27 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	121	147	370	80.26%
Function:	(215)	(46.64%)
Content:	(155)	(33.62%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(84)	(18.22%)
K2 Words (1001-2000):	15	15	21	4.56%
> Anglo-Sax:	(4)	(0.87%)
1k+2k	(84.82%)
AWL Words (academic):	17	19	26	5.64%
> Anglo-Sax:	(3)	(0.65%)
Off-List Words:	<u>2</u>	<u>33</u>	<u>44</u>	<u>9.54%</u>
	153+?	214	461	100%

Words in text (tokens):	461
Different words (types):	214
Type-token ratio:	0.46
Tokens per type:	2.15
Lex density (content words/total)	0.53

Pertaining to onlist only

Tokens:	417
Types:	181
Families:	153
Tokens per family:	2.73
Types per family:	1.18

Anglo-Sax Index: %

(A-Sax tokens + functors / onlist tokens)

Greco-Lat/Fr-Cognate Index: (Inverse of above) %

Current profile	
%	Cumul.
80.26	80.26
4.56	84.82
5.64	90.46
9.54	100.00

A. AWL Tokens lists

AWL [17:19:26] areas authorities circumstances consequence cultural culture debate debate environment establish estimated finally generation generation job legislation perceived prime prime prime removed removed removing style style style

Sublist 1

areas authorities environment establish estimated legislation

Sublist 2

consequence cultural culture finally perceived

Sublist 3

circumstances removed removed removing

Sublist 4

debate debate job

Sublist 5

generation generation prime prime prime style style style

B. AWL Types list

AWL types: [17:19:26] areas_[1] authorities_[1] circumstances_[1] consequence_[1] cultural_[1] culture_[1] debate_[2] environment_[1] establish_[1] estimated_[1] finally_[1] generation_[2] job_[1] legislation_[1] perceived_[1] prime_[3] removed_[2] removing_[1] style_[3]

C. AWL Families list

AWL families: [17:19:26]

area_[1] authority_[1] circumstance_[1] consequent_[1] culture_[2]
 debate_[2] environment_[1] establish_[1] estimate_[1] final_[1]
 generation_[2] job_[1] legislate_[1] perceive_[1] prime_[3] remove_[3]
 style_[3]

AWL Fr non-cognate families: [families 1 : tokens 3] remove_[3]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: Stolen Generation stunt (2,844 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): europeans, australia, aborigines, asian, australian, european, aboriginal, victorian, torres, strait, islander, john, howard, paul, keating, kevin, rudd, end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	132 (71.74)	159 (73.95)	390 (84.60)	84.60
K-2 Words :	24 (13.04)	26 (12.09)	39 (8.46)	93.06
K-3 Words :	18 (9.78)	18 (8.37)	21 (4.56)	97.62
K-4 Words :	4 (2.17)	4 (1.86)	4 (0.87)	98.49
K-5 Words :	3 (1.63)	3 (1.40)	3 (0.65)	99.14
K-6 Words :	1 (0.54)	1 (0.47)	1 (0.22)	99.36
K-7 Words :	1 (0.54)	1 (0.47)	1 (0.22)	99.58
K-8 Words :				
K-9 Words :	1 (0.54)	1 (0.47)	1 (0.22)	99.80
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				

K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	0 (0.00)	0 (0.00)	
Total (unrounded)	184+?	215 (100)	461 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	461
Different words (types):	215
Type-token ratio:	0.47
Tokens per type:	2.14

Pertaining to onlist only

Tokens:	461
Types:	215
Families:	184
Tokens per Family :	2.51
Types per Family :	1.17

Current profile (token %)	
K-1 (84.60)	84.60
K-2 (8.46)	93.06
K-3 (4.56)	97.62

K-4 (0.87)	98.49
K-5 (0.65)	99.14
K-6 (0.22)	99.36
K-7 (0.22)	99.58
K-9 (0.22)	99.80
OFF (0.00)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 109 : types 129 : tokens 365]

a_[8] act_[1] addition_[1] after_[1] agreed_[1] all_[1] already_[1] also_[1]
 an_[4] and_[10] are_[1] areas_[1] arrived_[1] as_[5] away_[5] be_[2]
 becoming_[1] been_[1] beginning_[1] better_[1] brought_[1] by_[2]
 cases_[1] chance_[1] child_[1] children_[8] church_[1] could_[1] country_[2]
 course_[1] did_[3] difficult_[2] do_[1] doing_[1] driven_[1] during_[1]
 earlier_[1] early_[1] excuse_[1] experienced_[1] families_[1] family_[2]
 few_[1] finally_[1] first_[3] for_[4] found_[1] from_[11] gave_[1] get_[1]
 give_[1] go_[1] government_[2] had_[3] happy_[1] have_[2] his_[1]
 home_[1] homes_[3] however_[1] in_[9] is_[4] islander_[1] islands_[1] it_[6]
 its_[1] job_[1] know_[1] land_[2] last_[1] led_[1] life_[3] made_[2] main_[1]
 make_[1] more_[2] most_[1] move_[1] nation_[1] natural_[1] necessary_[1]
 new_[1] north_[1] not_[5] number_[10] numbers_[3] of_[8] on_[3] one_[4]
 or_[2] own_[1] parents_[3] people_[3] place_[1] point_[1] proper_[1]
 protect_[2] protection_[1] race_[1] reached_[1] reasons_[1] right_[1]
 school_[1] settle_[2] settlers_[2] simple_[1] single_[1] some_[1] started_[2]
 starting_[1] stolen_[2] take_[2] taken_[3] than_[2] that_[2] the_[38]
 their_[10] them_[2] there_[1] these_[7] they_[5] think_[1] this_[3]
 thousands_[1] time_[1] to_[17] trying_[1] under_[1] understand_[1]
 unknown_[1] until_[1] very_[1] wanted_[1] was_[7] were_[10] when_[2]
 which_[1] white_[1] whites_[1] who_[2] why_[1] with_[1] years_[5] you_[1]

BNC-COCA-2,000 types: [fams 24 : types 26 : tokens 39]

circumstances_[1] cultural_[1] culture_[1] demands_[1] diseases_[1]
 environment_[1] establish_[1] february_[1] generation_[2] minister_[3]
 native_[1] neither_[1] nor_[1] official_[5] opportunities_[1] population_[1]
 prime_[3] referred_[1] regarded_[1] removed_[2] removing_[1] shores_[1]
 spared_[1] style_[3] sufferings_[1] threat_[2]

BNC-COCA-3,000 types: [fams 18 : types 18 : tokens 21]

abuse_[1] apology_[2] approve_[1] authorities_[1] consequence_[1]
 continent_[2] debate_[2] estimated_[1] fertile_[1] foster_[1] inhabited_[1]

legislation_[1] neglect_[1] perceived_[1] permanent_[1] religions_[1]
 rituals_[1] shortly_[1]

BNC-COCA-4,000 types: [fams 4 : types 4 : tokens 4]

cultivate_[1] indigenous_[1] mid_[1] predecessor_[1]

BNC-COCA-5,000 types: [fams 3 : types 3 : tokens 3]

descent_[1] inferior_[1] orphanages_[1]

BNC-COCA-6,000 types: [fams 2 : types 2 : tokens 2]

halved_[1] strait_[1]

BNC-COCA-7,000 types: [fams 2 : types 3 : tokens 9]

aboriginal_[5] aborigines_[3] nomadic_[1]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

smallpox_[1]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

B. Families list

BNC-COCA-1,000 Families: [fams 109 : types 129 : tokens 365]

a_[12] act_[1] add_[1] after_[1] agree_[1] all_[1] already_[1] also_[1]
and_[10] area_[1] arrive_[1] as_[5] away_[5] be_[25] become_[1] begin_[1]
better_[1] bring_[1] by_[2] case_[1] chance_[1] child_[9] church_[1]
could_[1] country_[2] course_[1] difficult_[2] do_[5] drive_[1] during_[1]
early_[2] excuse_[1] experience_[1] family_[3] few_[1] final_[1] find_[1]
first_[3] for_[4] from_[11] get_[1] give_[2] go_[1] govern_[2] happy_[1]
have_[5] he_[1] home_[4] however_[1] in_[9] island_[2] it_[7] job_[1]
know_[2] land_[2] last_[1] lead_[1] life_[3] main_[1] make_[3] more_[2]
most_[1] move_[1] nation_[1] nature_[1] necessary_[1] new_[1] north_[1]
not_[5] number_[13] of_[8] on_[3] one_[4] or_[2] own_[1] parent_[3]
people_[3] place_[1] point_[1] proper_[1] protect_[3] race_[1] reach_[1]
reason_[1] right_[1] school_[1] settle_[4] simple_[1] single_[1] some_[1]
start_[3] steal_[2] take_[5] than_[2] that_[2] the_[38] there_[1] they_[17]
think_[1] this_[10] thousand_[1] time_[1] to_[17] try_[1] under_[1]
understand_[1] until_[1] very_[1] want_[1] when_[2] which_[1] white_[2]
who_[2] why_[1] with_[1] year_[5] you_[1]

BNC-COCA-2,000 Families: [fams 24 : types 26 : tokens 39]

circumstance_[1] culture_[2] demand_[1] disease_[1] environment_[1]
establish_[1] february_[1] generation_[2] minister_[3] native_[1] neither_[1]
nor_[1] official_[5] opportunity_[1] population_[1] prime_[3] refer_[1]
regard_[1] remove_[3] shore_[1] spare_[1] style_[3] suffer_[1] threat_[2]

BNC-COCA-3,000 Families: [fams 18 : types 18 : tokens 21]

abuse_[1] apology_[2] approve_[1] authority_[1] consequence_[1]
continent_[2] debate_[2] estimate_[1] fertile_[1] foster_[1] inhabit_[1]

legislate_[1] neglect_[1] perceive_[1] permanent_[1] religion_[1] ritual_[1]
shortly_[1]

BNC-COCA-4,000 Families: [fams 4 : types 4 : tokens 4]

cultivate_[1] indigenous_[1] mid_[1] predecessor_[1]

BNC-COCA-5,000 Families: [fams 3 : types 3 : tokens 3]

descent_[1] inferior_[1] orphan_[1]

BNC-COCA-6,000 Families: [fams 2 : types 2 : tokens 2]

halve_[1] strait_[1]

BNC-COCA-7,000 Families: [fams 2 : types 3 : tokens 9]

aborigine_[8] nomad_[1]

BNC-COCA-8,000 Families: [fams : types : tokens]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 1]

smallpox_[1]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.2.2.7 Effects of Removal

7.2.2.7.1 Text only file

The Effects of Removal on American Indian Tribes

Clara Sue Kidwell , University of North Carolina at Chapel Hill

National Humanities Center

Echohawk / Gilcrease

Brummett Echohawk (Pawnee artist) , " Trail of Tears , " 1957

The removal of American Indian tribes from lands east of the Mississippi River to what is now the state of Oklahoma is one of the tragic episodes in American history . Early treaties signed by American agents and representatives of Indian tribes guaranteed peace and the integrity of Indian territories , primarily to assure that the lucrative fur trade would continue without interruption . American settlers' hunger for Indian land , however , led to violent conflict in many cases , and succeeding treaties generally compelled tribes to cede large areas to the United States government .

Howard

Norma Howard (Choctaw artist) , " Choctaw Village , " ca. 1998

One rationale for these treaties was that Indians were migratory hunters who only followed the game and had no attachment to any particular lands . This rationale ignored the fact that tribes in the south east raised significant crops of corn and lived in settled villages . Americans were already swayed by arguments based on stereotypes of Indians as hostile , savage , wandering people . For students , the question is to what extent these stereotypes still persist in their thinking .

Orders for removal of Cherokee from North Carolina , Georgia , Tennessee and Alabama , 1838

Orders for removal of Cherokee from Georgia , 1838

Notice to the Cherokee that steam boats will be available for their transportation to territory west of the Mississippi , 1838

The final removal came under the Indian Removal Act . Missionary societies who had invested their time and money teaching Indians to live with their white neighbors and accept Christianity lobbied Congress to oppose the act . It finally passed , but only by a one vote margin , in September of 1830 . The Choctaw , Cherokee , Chickasaw , Creeks , and Seminoles signed treaties agreeing to leave their homes in the south east and move west . Their travels were marked by outbreaks of cholera , inadequate supplies , bitter cold , and death from starvation and exhaustion . The Cherokees' march was a forced one under the direction of the United States army , and it came to be known as the Trail of Tears or , in their own term , The Place Where They Cried . Removal was a tragedy as thousands of people were forced to leave behind their homes , live stock , crops , and places that had spiritual significance for them .

McClung Museum

Depiction of Cherokee farmstead of the mid 1700s , based on historic descriptions and archaeological excavations from the Lower Little Tennessee River Valley , eastern Tennessee , 1967 - 1982

Archaeological evidence , native oral traditions , and written sources help us reconstruct the past and understand the way in which landscape shaped the culture and the history of these people in their original home lands and how they had to adapt to a new environment west of the Mississippi River .

The Choctaw , Chickasaw , Cherokee , Creek , and Seminole tribes lived originally in the area that now encompasses the states of Mississippi , Alabama , Georgia , Tennessee , and North Carolina . These groups defined their own identity in many ways , but an important one was their relationship with the land that they considered their home .

The Choctaw territory in present day Mississippi extended from the Mississippi Delta on the west , through rich , black soil prairie lands in the north east , to piney woods in the southern part of the state . Its eastern boundary was defined by the watershed of the Black Warrior River , and the Pearl , Tombigbee , and Chickasawhay Rivers defined its three major divisions — the Okla Falaya , the Okla Tanap , and the Okla Hannali (Okla being the Choctaw word for " people ") .

Tribal regions before Removal , ca. 1830

See descriptions of the tribal regions .

The Creeks lived in Alabama and south western Georgia —

the " upper " Creeks along the Tallapoosa and Coosa Rivers and the " lower " Creeks along the Chattahoochee River .

The Chickasaw home land was in the upper Mississippi Delta region in northern Mississippi , into western Tennessee and northern Alabama .

The Cherokees occupied the valleys of the southern Appalachian Mountains , establishing villages along the Tennessee River and its tributaries . They included five divisions (as defined by the British colonial government in the 1700s) : the Lower Towns in north Georgia , the Over-the-Hill (or Overhill) Towns in eastern Tennessee , and Middle Towns , Valley Towns , and Out Towns in western North Carolina .

The Seminoles , originally of Lower Creek identity , emerged as a distinctive tribal group in

the early to mid 1700s as a result of conflict between European colonists and tribal villages . A major uprising by tribes along the east coast of Georgia , the Yamasee Rebellion of 1715 , led to military action on the part of the British that destroyed native villages and dispossessed their populations . Homeless groups moved south into Spanish Territory below the 31 parallel (which became Florida) , as the Spanish were reputed to have a liberal policy toward Indians and to leave them in peace .

The Indian groups that settled in what is now Florida and the southern portions of Georgia , Alabama , and Mississippi , came to be called Seminoles , a corruption of the Spanish term " cimimerones " or wanderers .

Each of these tribal groups had its own origin tradition .

Winston County

Nanah Waiya Indian mound , Mississippi

The Choctaws and Chickasaws shared a common origin tradition , that they had lived west of the Mississippi River and had migrated to the east . The migration was the result of the dream of a holy man that the sacred pole that stood in the center of his village would lean in the direction of the march . It was led by two brothers , Chata and Chiksa . During the long journey and after the people crossed the Mississippi , the brothers and their followers were separated by disagreement , in a thunder storm , the accounts vary . Chata and his people followed the pole until it finally stood upright near a hill . The site today is at Nanah Waiya , a flat topped mound about twenty miles north of Philadelphia , Mississippi , the tribal headquarters of the Mississippi Band of Choctaw Indians .

North Georgia mountains

The Cherokee origin tradition explains the formation of their home land — the hills and valleys of the southern Appalachian Mountains , along the Tennessee River and its tributaries . When the earth was created and the land was very soft , birds were sent down from the sky to find a dry place for the animals to live . When they were unsuccessful , a giant buzzard was sent to continue the search . As he grew tired he flew lower and lower , and his wing tips began to hit the soft new land , pushing down the valleys and raising the hills .

Duncan

Etowah River , near the Coosa River , Alabama

The Creeks occupied villages along the Chattahoochee , Tallapoosa , and Coosa Rivers in Alabama . Their origins began under a mountain in the west , which opened up and the people emerged and settled nearby . But the earth opened up and ate their children , and they began a long march to the east , crossing several rivers . They encountered three other peoples , from whom they learned the use of certain herbs . They also found a pole on a mountain , which became their guide . They finally encountered a white path , which they followed to Caloose Creek . They found the people who had made the white path and settled near them . The story explains not only how the Creeks came to the south east but how they came to dominate most of what is now the state of Alabama by making alliances with tribal groups whose hunting territories they wanted . The Creeks were a confederacy of peoples held together by similarity of language .

National Archives

Bankhead National Forest , Alabama

" Their environments shaped their senses of identity . "

These five tribes of the south east were village dwellers . They clustered around streams and rivers , which generally defined territorial hunting ranges . They raised numerous varieties of corn , beans , and squashes , but their primary supply of meat came from hunting . Deer , bear , and wood land buffalo were their prey .

Their environments shaped their senses of identity . The tribes of the south east maintained a delicate balance with the forces of the environment around them . The woods were full of spiritual forces who could harm someone who wandered alone into their domain . Violent storms , sudden floods in the river valleys , lightning set fires in the woods , all were reminders of the power of the world . The Green Corn ceremony , variations of which occurred in the Cherokee , Creek , and Chickasaw communities , renewed the world in the spring for the upcoming year .

During the late eighteenth century , major changes began to affect the life styles of the south eastern native people . The introduction of domesticated live stock among the Choctaws in the 1790s provided a new source of food that began to replace deer meat in the diet . Hunting deer for skins to trade with French and English agents had depleted deer populations throughout the south east . Although domesticated cattle roamed free in the forests and prairies , they could be easily captured . Other introductions to the Choctaw diet included domesticated pigs and potatoes , and some families cultivated fields of cotton . By the early 1800s a missionary could report that Choctaw women had spinning wheels , cards , and were weaving yards of cloth .

Voluntary removal , late 1700s - early 1800s

although Indian removal is generally associated with the 1830 act of Congress , the process was already beginning by the late 1700s . Pressure of white settlement led small parties of Choctaws , Cherokees , and Chickasaws to move west of the Mississippi , and by 1807 they were settling in Arkansas , Indian Territory , and east Texas . There they could hunt and raise their crops . This voluntary removal to escape conflict with white settlers and government agents thus preceded forced removals .

Federal policy toward Indians was ambivalent . Thomas Jefferson acquired the Louisiana Territory in part to find a place for Indian communities who would not assimilate into white society and who wished to pursue their traditional hunting ways of life , but he also promoted government run trading posts in Indian country so that Indians would build up such great debts that they would be willing to give up some of their land in payment . Indians might choose to move , but Jefferson also found ways to force them to make the choice .

Piereman

Warrant issued to a Revolutionary War soldier for 100 acres of western land as payment for his service , 1784

Despite the integration of domesticated cattle and the technology of weaving into their life styles , Americans still considered the south eastern tribes savages . The increasing American

population led to pressure to develop new western lands . The War of 1812 , a definitive victory over the English , gave Americans a sense of national identity , but it also created a need for Indian land . The United States paid its soldiers from the Revolutionary War and the War of 1812 not with money but with warrants that they could exchange for western land .

" In going up the stream there were houses and farms on both banks of the River . The houses were decently furnished , and their farms were well fenced and stocked with cattle . They had everything they needed : food , clothes , water and good land . "

Nuttall , Journal , 1819 , on a Cherokee band in the Arkansas Territory

The pressure for the development of western lands required the removal of Indians from those lands . Even while government agents were holding out promises of western lands that would be theirs forever , Americans were exploring those lands . In 1819 , Thomas Nuttall , an English botanist , traveled to the Arkansas territory . His account painted a picture of a fertile and productive environment for agriculture , a description seemingly designed to inspire interest in the minds of land speculators . The Choctaw leader Pushmataha , however , when pressed to sign a treaty ceding his tribe's land in central Mississippi in exchange for others in the west , protested : " We wish to remain here , where we have grown up as the herbs of the woods ; and do not wish to be transplanted into another soil . "

" Indeed most of the streams on this side of the Arkansas are said to afford springs of salt water which might be wrought with profit."

Thomas Nuttall , A Journal of Travels into the Arkansas Territory during the Year 1819
In the period between 1817 and 1825 , however , the tribes signed treaties agreeing to exchange eastern lands for western ones . These early treaties did not require the tribes to move west , and most remained in their homes , but small vanguards crossed the Mississippi to take up residence in the new territory , some joining relatives already settled there . Some Choctaw families moved after the Treaty of Doaks Stand , signed in 1820 . Some Creek and Cherokee groups moved west after treaties they signed in 1818 .

The pressures on the tribes culminated in 1829 and 1830 when the legislatures of Mississippi and Georgia passed laws to extend their jurisdiction over the Choctaw , Chickasaw , and Cherokee Nations . The actions brought into sharp relief the dilemma that faced the tribes . Were they to submit to the laws of a foreign government to remain in the lands that they considered their home land , or were they to move to the west to retain their autonomy ?

Sorrows of the Seminoles , Banished from Florida , ca. 1835

Song about the Seminoles' departure sung in

the Muskogean language . Library of Congress

Congress followed the actions of the states with the 1830 Indian Removal Act that directed the federal government to negotiate with Indian tribes to exchange their lands east of the Mississippi River for lands to the west . Under the provisions of the act , the Choctaws , the Chickasaws , Cherokees , Creeks , and ultimately the Seminoles , who had fled to Florida in the early nineteenth century , moved to Indian Territory (what is now the state of Oklahoma) in the period from 1831 through the 1840s .

it is (with sorrow) that we are forced by the authority of the white man to quit the scenes of our childhood , but stern necessity says we must go , and we bid a final farewell to it and all we hold dear East of the Father of Waters . . . "

George Hicks , Cherokee , on the " Trail of Tears , " November 1838

Forced removal , 1830s - 1850 large map

Thus the five tribes moved under duress to the Indian Territory of eastern Oklahoma . The story of their hardships on the journey is well known . Here we consider another aspect of their experience—the new environments to which they had to adapt , and adapt quickly . I recommend that you study the maps of Indian Territory below before continuing your reading . Visualize the areas they left — their home lands east of the Mississippi—and the new lands west of the Mississippi to which they were forcibly removed . Compare the physical aspects of the regions they left and of the regions they settled . (For more detailed ecological comparisons, see the Physical Environment links in the online resources.)

Library of Congress

1892

Oklahoma (shaded relief)

Tribal regions in Indian Territory after Removal

See descriptions of the regions .

The Cherokees settled in the north east of the new territory . Their homes in the Appalachians had been dominated by mountain ranges , rivers , and forests . In the foothills of the Ozark Mountains and the valleys of the Illinois , Arkansas , Grand , and Verdigris rivers they found lands similar to what they had known before , but foreign because they were brought there under duress . Of the five tribes , the Cherokee suffered probably the harshest conditions during their removal . In the south east , they had lived in villages along river valleys where they planted their crops on river terraces and hunted over large areas . In their first year in the west , they planted along the Arkansas River , which flooded , as it did regularly , and the first crops were washed out .

The Creek people settled in the central part of the Indian Territory . The northern and southern branches of the Canadian River bounded their territory , and numerous creeks fed into those rivers . The low hills and a narrow band of dense forest known as the Cross Timbers distinguished the area .

The Choctaws moved into the area of the San Bois and Ouachita mountains and the Kiamechi river in the south east region of the territory . The piney woods , mountains , and rivers of the region were similar to those of the south eastern area in Mississippi . Although the topography was familiar , the Choctaws had had to leave behind their homes , fields , crops , and whatever live stock they possessed .

The Chickasaws moved into Choctaw territory in 1837 with the promise that they would occupy its western portion , the land between the Cross Timbers and the open space of the Plains . Because the land in what was known as the Indian Territory had been assigned to the Creeks , Cherokees , and Choctaws , there was no place for the Chickasaws . They had sold their eastern lands to the United States government for approximately \$ 500 ,000 , with which they could buy a new home land . With this money , they leased land from the Choctaws . The money also created a trust fund that yielded an income for the tribe of

between \$ 60,000 and \$ 75,000 a year . They could live on annuity payments without having to establish farms . For the Chickasaws , removal led them into a cash economy and a political situation that stifled their dependency upon the natural environment .

The Seminoles resisted removal in a series of hard fought and costly wars from the 1810s to the 1850s . In 1835 , about 4,000 Seminoles were captured and sent to the Indian Territory , where they were located in the western section of the Creek territory . Another small group was sent from Florida in the late 1850s , when the government campaign to remove the south eastern Indians came to an end .

Van Horn

North east Alabama
Lewallen

South east Oklahoma

" the hilly , wooded south eastern part of the Territory that resembled their home lands in the south east [US] "

As the tribes entered their new lands , the one thing they would not do was move beyond the hilly , wooded south eastern part of the Territory that resembled their home lands in the south east . Further west , the dramatic opening of the Great Plains with its vast , treeless , arid expanses of territory , was foreign to their experience . In addition , it was dominated by Kiowas, Comanches, Wichitas, and Apaches buffalo hunting , highly mobile societies whose raids were a threat to the settled villages of the South eastern tribes . Although their treaties guaranteed their rights to lands all the way to the head waters of the Arkansas , Red , and Canadian Rivers , the environment in the west created a natural boundary beyond which the south eastern tribes would not move .

Denver Public Library

Plains in south western Oklahoma near Fort Sill , with Kiowa or Apache camp , c. 1875
" the environment in the west created a natural boundary beyond

which the south eastern tribes would not move "

Although the terrain was different , one element of native knowledge that persisted and adapted from the south east to the Indian Territory was the use of herbal medicines . In the west Choctaws in the early 1900s century used a tea made from boiled black root as a laxative , blood weed for purifying blood , black root and fall willow for measles and smallpox (European introduced diseases) , and broom weed for colds and coughs . It could also prevent pneumonia if taken in time . Other medicines described by Choctaws in Oklahoma include Sycamore bark , which was boiled into a tea for coughs, slippery elm , which , was mixed with milk and used for burns , and " rusty water " — water in which iron chains were allowed to stand for a few days — which was used as a tonic . The use of rusty water was obviously an adaptation to white society .

McClung Museum

Collection of Cherokee herbal medicines , ca. 1827

Place cursor on recipes for text .

The use of plants for medicine was not unique to native people , either in the south east or in the Indian Territory . European settlers had brought well established beliefs in the power of herbal medicines , which were based on similarities of form between plants and the human body , while Native beliefs were based on the idea that plants were living beings . Choctaw people adapted the plants of their new environment to their beliefs in herbal medicines .

" Within the past six years , the Indian's sentiments have undergone a radical change respecting railroads . He now hauls to the stations on the line his pecans , pork , cotton , and his surplus game , receives a liberal sum of money in exchange, and goes home satisfied that the railroad is a friendly institution . "

Jenness , " The Indian Territory , " The Atlantic Monthly , April 1879

As the tribes of the south east moved to new lands in the west , they had already entered an economic system that made land a commodity with a monetary value . Domesticated animals horses , cattle , pigs - had replaced the game that they had hunted to supply the earlier European trade in meat and hides . In 1840 a missionary described the Choctaws in Indian territory as living in log cabins , raising corn , pumpkins , peas , melons , and yams . Their farms generally ranged from one to ten acres , and black slaves were generally used as field hands on larger farms . Men worked the fields , and hunting was limited to small animals such as rabbits and squirrels to supplement the family diet . A favorite settling place for wealthier tribal members was by

Sandra Riley

Choctaw coal mine , Lehigh (Coal County) , Oklahoma , ca. 1920

water falls that would run their grist mills for their grain . Coal and oil deposits in the Choctaw territory provided a new source of wealth in the later part of the century , and railroads which began to cross Indian territory after the Civil War led to a demand for timber for railroad ties and stone .

Through the process of removal , Indians had to adapt to both new environments and a new sense of their place in American society . The tribes of the south east adapted to a new environment , but one that , like America in general , was exploiting natural resources for economic development . The forces at work in America in the latter part of the nineteenth Valjean Hessing/Heard Museum

Valjean Hessing (Choctaw artist) , Choctaw Immigrants , 1972

It would not , however , totally destroy their religious connection to it . Cherokee and Creek ceremonial grounds persist in Oklahoma today . Some Choctaws still use herbal medicines . The origin traditions that explained their original home lands are preserved in written form and in stories that may still be told in some homes . The history of their removal is also recorded in books and stories passed down through generations . The history of removal is part of the identity of members of the Choctaw , Chickasaw , Cherokee , Creek and Seminole tribes . It is an essential part of explaining the role of changing environments for contemporary tribal members .

SCHOLARS DEBATE

Native American Images

(Cherokee artist) , Men with Broken Hearts , 1994

The Trail of Tears has become the symbol in American history that signifies the callousness of American policy makers toward American Indians . Indian lands were held hostage by the states and the federal government , and Indians had to agree to removal to preserve their identity as tribes .

The factors leading to Indian removal are more complex . Early writers such as Annie Heloise Abel and Grant Foreman simply described the policy and events . Foreman's book , Indian Removal (1932) , is compelling because the reader can draw from quotes from primary documents the details of the removal experience for the five south eastern tribes . The bulk of the literature on removal deals with the impact on the Choctaws , Cherokees , Chickasaws , Creeks , and Seminoles , but Abel's work , Events Leading to the Consolidation of American Indian Tribes West of the Mississippi River (1906) deals with the wider implications of the policy for other tribes in other parts of America .

The complexity of reasons for removal comes from later historical interpretation . Richard White's The Roots of Dependency (1983) puts the Choctaws in the larger context of American history and explains their experience in light of the changing economy of American society in the post revolutionary war era . The religious justification for removal , preservation of Indian nations from the pernicious influence of white populations , is apparent in George Schultz's An Indian Canaan (1972) , the story of Isaac McCoy , the Baptist missionary who was the most active proponent of an Indian state , where Native peoples could be consolidated in an area where , if the environment was foreign , they could be protected to pursue their own life style .

" . . . we have done so much to destroy the Indians , and so little to save them ; and that , before another step is taken , there should be the most thorough deliberation , on the part of all our constituted authorities , lest we act in such a manner as to expose ourselves to the judgments of heaven . "

Jeremiah Evarts , Essays on the present crisis in the condition of the American Indians , 1829

The moral objections to removal are evident in the writings of Jeremiah Evarts , Secretary of the American Board of Commissioners for Foreign Missions , the organization that established the first Christian missions among the Cherokees and Choctaws in the early 1800s . Cherokee Removal : The " William Penn " Essays and Other Writings (1981) is a collection of Evarts's letters and essays . Evarts upheld an inherent right of Native people to be secure in their lands . His covert agenda was to protect the financial investment that the American Board had made in the mission buildings that they had established in the south east .

The impact of removal on native populations has led to some debate in terms of demographics. The extent of the loss of life among migrants has an impact on the ability of people to maintain community structures such as clan and kin relationships. Loss of large numbers of family members through epidemic disease and the rigors of removal disrupt communities. Debates about the impact of epidemic disease and depopulation continue among scholars today. For the Cherokee Trail of Tears, consult Russell Thornton's The Cherokees: A

Population History (1990), in which he estimates both loss of life and the potential population of the Cherokee nation had Removal not taken place .

Census of Cherokee families in Georgia , North Carolina , and Tennessee (probably the 1840 federal census) ; excerpt . Phrase " died during the emigration " appears repeatedly in the remarks . University of Georgia Libraries .

Place cursor on " Remarks " entries for transcribed text .

National Archives

Delegates from 34 tribes in front of Creek Council House , Indian Territory , 1880

The dynamic ability of tribes to adapt to new environments is evident in William McLoughlin's After the Trail of Tears : The Cherokees Struggle for Sovereignty 1839 - 1880 (1993) . Although the usual historical interpretation of the Trail of Tears has portrayed Indians as victims of federal policy , renewed attention to earlier scholarship such as Grant Foreman's works shows that Indians were making decisions to move west of the Mississippi long before the Removal Act . Those decisions may have some basis in traditions that they had originally lived west of the Mississippi . The historical tragedy and loss of home lands has been emphasized . The resilience of tribes and their ability to adapt to new environments needs to be stressed . In the larger scheme of American history , many tribal members were adapting to a new kind of economic system as were Americans generally . They faced the pressures of a market economy in which land was becoming a commodity to be bought and sold .

The result was a historical experience that for contemporary tribal members joins traditional origin stories with accounts of the experiences of their ancestors in moving to and adapting to a new environment .

Library of Congress

" many tribal members were adapting to a new kind of economic system as were Americans generally "

GUIDING STUDENT DISCUSSION

Students should consider the factors that motivated the United States government to attempt to dispossess whole groups of peoples of their lands . In a comparative sense , the Holocaust in Europe , the internment of Japanese during World War , and , today , ethnic violence in Bosnia , Africa , and Chiapas , Mexico , can show that the motives of nationalism and " ethnic cleansing " still go on in the modern world .

European colonialism brought people to the new world to exploit its natural resources . Native people lived in an environment which they considered as their sources of power . Ceremonies such as the Green Corn Dance were part of the process by which they interacted with the environment , and they believed that they played a causal role in the processes of the environment . (Christian churches had replaced traditional ceremonialism for many tribal

members , but ceremonies such as the Green Corn dance persisted .) An origin story from a local tribe , and a description of a ceremony , can show how Indian people explained the form of their environment , the spiritual beings who were involved in its creation , and the way in which ceremonies play a role in recreating the world (see online resources) . Students should think about Christian attitudes toward nature — are humans indeed stewards of the land ? Should they feel a responsibility to the land ? How do they explain their own environments in a scientific sense , but also , what are their particular memories of significant places in their own lives ? What does the notion of place convey to them ?

Motto : " Work Conquers All"

From the Native perspective , the choice can be presented simply as that between remaining a self governing people with their own laws in a new land , or remaining in their homes as subjects of a foreign government . This simple dichotomy does not mean that Native communities were unified in making one choice or the other . What choices would students make in those circumstances , and why ? Many Native leaders were confronted with the necessity of resisting pressures for removal when they personally favored the move to protect tribal sovereignty .

Students should think about the choices that Europeans made to seek religious freedom in new lands, and how they moved into what they thought was virgin territory. Who has rights to the environment, and for what reasons?

American nationalism is an important element in understanding Indian removal . Indian nations were the original inhabitants of the land , and they are designated specifically in the Constitution . The fact of Indians as sovereign nations at the time of contact is essential to an understanding of American history . Students should be challenged to clarify their views of the extent and power of the original American government and the struggle to form an appropriate government for thirteen very disparate groups of people . They should also understand the fear that Americans felt about the continuing threat of British influence over American Indians during and after the Revolutionary War and up to the War of 1812 . Indian captivity narratives , which are available in various collections , can convey the factors in this fear . Biographies of Tecumseh will , however , present a Native view that Americans were a threat to Indian cultures and tribal integrity .

of presidents' State of the Union addresses

Arguments for removal by government officials show the reasons behind removal policy .

The annual addresses to Congress of presidents Madison , Monroe , and Jackson present the humanitarian view that Indians will perish under the superiority of American society unless they are removed from contact with that society . Within those addresses , there is also the idea that Indians do not utilize the land to its full productivity , and they should give way to American agriculturalists . At this point , students can consider that the traditional lives of the south eastern tribes who were the major objects of removal policy were indeed those of village farmers . Students should be challenged to reconcile these traditional life styles with the words of government officials .

The point of this lesson with regard to American environmental history should be the ways in which different groups of people see their relationship to their environments and how the conflict of different ideas has led to removal of people to seek freedom in new environments .

Clara Sue Kidwell is currently Director of the American Indian Center at the University of North Carolina at Chapel Hill . Her tribal affiliations are Choctaw and Chippewa . She was born in Tahlequah , Oklahoma , and raised in Muskogee , Oklahoma . She received a B.A. in Letters (1962) and a M.A. and Ph.D. in History of Science (1970) from the University of Oklahoma . Before joining the faculty there in 1995 she served for two years as Assistant Director of Cultural Resources at the National Museum of the American Indian , Smithsonian Institution . Her previous teaching positions include : Associate Professor and Professor of Native American Studies at the University of California at Berkeley (1974 - 95) ; Visiting Assistant Professor in Native American Studies at Dartmouth College , Hanover , New Hampshire (1980) ; Assistant Professor of American Indian Studies at the University of Minnesota (1972 - 74) ; Instructor of Social Sciences at Haskell Indian Junior College in Lawrence , Kansas (1970 - 72) ; and Instructor at the Kansas City Art Institute (1968 - 69) .

She has taught courses on American Indian history , philosophy , and medicine and has published a number of articles including " Systems of Knowledge , " in *America in 1492* , edited by Alvin Josephy (Knopf , 1991) ; " Indian Women as Cultural Mediators , " *Ethno history* 39 : 2 (Spring 1992) , 97 - 107 ; " Choctaw Women and Cultural Persistence in Mississippi , " in *Negotiators of Change : Historical Perspectives on Native American Women* , edited by Nancy Shoemaker (Routledge , 1995) ; and *Choctaws and Missionaries in Mississippi , 1818 - 1918* (University of Oklahoma Press , 1995) .

7.2.2.7.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. **Contractions that are written out:**

3. **Hyphenated words with hyphen removed:**

one-vote, present-day, Over-the-Hill, flat-topped, lightning-set, government-run, hard-fought, Apaches—buffalo-hunting, European-introduced, well-established, post-revolutionary, self-governing,

4. **Compound words separated:**

southeast, **southeastern**,

5. **Words (groups of letters) removed from the text analysis:**

UGA Lib. (2), Tenn. SL&A, GA DITT, FLA SA, Sterner/JHU, – references to sources

“st”- 31st

“IP” world war II

Note: To lessen the load of proper nouns in this text analysis the pictures and captions related to these have been removed from the text analysis. This has been done after the heading “Guiding Student Discussion”.

6. **Proper nouns:**

Alabama, American, Americans, Appalachian, Bankhead, British, Brummett, Carolina, Caloose, Chata, Chattahoochie, Cherokee, Cherokees, Chickasaw, Chickasaws,

Chickasawhay, Chiksa, Choctaw, Choctaws, Cimmerones, Clara, Coosa, Coosa, Creek, Creeks, Duncan, Echohawk, English, Etowah, European, Falaya, Florida, French, Georgia, Gilcrease, Hannali, Howard, Indian, Indians, Jefferson, Kidwell, Louisiana, McClung, Mississippi, Nanih, Norma, Nuttall, Okla, Oklahoma, Overhill, Pawnee, Philadelphia, Piereman, Pushmataha, Seminoles', Seminoles, September, Spanish, Sue, Tallapoosa, Tanap, Tennessee, Texas, Thomas, Tombigbee, Waiya, Winston, Yamasee, Doaks, Muskogee, Oklahoma, George, Hicks, Appalachians, Ozark, Arkansas, Illinois, Verdigris, Canadian, San, Bois, Ouachita, Kiamechi, Van, Horn, Lewallen, Kiowas, Comanches, Wichitas, Apaches, Sill, Kiowa, Sycamore, Jenness, Sandra, Riley, Lehigh, Valjean, Hessing, Donald, Vann, Annie, Heloise, Abel, Grant, Foreman, Richard, White's, George, Schultz's, Isaac, McCoy, Baptist, Jeremiah, Evarts, Christian, William, Penn, Georgia, William, McLoughlin's, Grant, Foreman's, Holocaust, Japanese, Tecumseh, Madison, Monroe, Jackson, Tahlequah, B.A., M.A., Ph.D., Smithsonian, Native (American) Studies, California, Berkeley, Dartmouth, Hanover, Hampshire, Minnesota, Haskell, Lawrence, Kansas, Alvin, Josephy, Knopf, Nancy, Shoemaker, Routledge, Christianity, McClung, Russell, Thornton, Denver, Bosnia, Africa, Europeans, Mexico, Chiapas, April, November, Chippewa,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

University of North (Carolina), Chapel Hill National Humanities Center, Trail of Tears, (Mississippi) River, United States, (Indian) Removal Act, Christianity, Congress, The Place Where They Cried, (McClung) Museum, Lower Little (Tennessee) River Valley, (Mississippi) Delta, Black Warrior River, Pearl River, (Tombigbee) River, (Chickasawhay) River, (Tallapoosa) River, (Coosa) River, Chattahoochie River, (Appalachian) Mountains, (Tennessee) River, Lower Towns, (Overhill) Towns, Middle Towns, Valley Towns, Out Towns, Lower (Creek), (Yamasee) Rebellion, (Winston) County, (Nanih Waiya Indian) mound, (Etowah) River, (Coosa) River, National Archives, (Bankhead) National, Forest, Green Corn ceremony, (Indian) Territory, (Louisiana) Territory, Revolutionary War, War of 1812, (Nuttall) Journal, Treaty of (Doaks) Stand, Library of Congress, Father of Waters, (Ozark) Mountains, (Illinois) River, (Arkansas) River, Grand River, and (Verdigris) River, (Canadian) River, University of (Georgia) Libraries, Cross Timbers, (San Bois) Mountains, (Ouachita) mountains, (Creek) Council House, (Kiamechi) River, Great Plains, Denver Public Library, Fort (Sill), (McClung) Museum Collection, The Atlantic Monthly, Coal County, Civil War, World War II, Bosnia, Africa, Chiapas, Mexico, Constitution, University of (Oklahoma), National Museum of the (American Indian), (Smithsonian) Institution, Native (American) Studies, University of (California), (Dartmouth) College, (New) Hampshire, University of (Minnesota), Haskell Indian Junior College, Lawrence, Kansas, Kansas City Art Institute, (Caloose) Creeek,

Note: Text related to illustrations have been included in the text analysis.

7.2.2.7.3 Text analysis

1. VP-Classic

Words recategorized by user as 1k items (proper nouns etc): ALABAMA, AMERICAN, AMERICANS, APPALACHIAN, BANKHEAD, BRITISH, BRUMMETT, CAROLINA, CALOOSE, CHATA, CHATTAHOOCHEE, CHEROKEE, CHEROKEES, CHICKASAW, CHICKASAWS, CHICKASAWHAY, CHIKSA, CHOCTAW, CHOCTAWS, CIMMERONES, CLARA, COOSA, COOSA,

CREEK, CREEKS, DUNCAN, ECHOHAWK, ENGLISH, ETOWAH, EUROPEAN, FALAYA, FLORIDA, FRENCH, GEORGIA, GILCREASE, HANNALI, HOWARD, INDIAN, INDIANS, JEFFERSON, KIDWELL, LOUISIANA, MCCLUNG, MISSISSIPPI, NANIH, NORMA, NUTTALL, OKLA, OKLAHOMA, OVERHILL, PAWNEE, PHILADELPHIA, PIEREMAN, PUSHMATAHA, SEMINOLES, SEMINOLES, SEPTEMBER, SPANISH, SUE, TALLAPOOSA, TANAP, TENNESSEE, TEXAS, THOMAS, TOMBIGBEE, WAIYA, WINSTON, YAMASEE, DOAKS, MUSKOGEE, OKLAHOMA, GEORGE, HICKS, APPALACHIANS, OZARK, ARKANSAS, ILLINOIS, VERDIGRIS, CANADIAN, SAN, BOIS, OUACHITA, KIAMECHI, VAN, HORN, LEWALLEN, KIWAS, COMANCHES, WICHITAS, APACHES, SILL, KIOWA, SYCAMORE, JENNESS, SANDRA, RILEY, LEHIGH, VALJEAN, HESSING, DONALD, VANN, ANNIE, HELOISE, ABEL, GRANT, FOREMAN, RICHARD, WHITE'S, GEORGE, SCHULTZ'S, ISAAC, MCCOY, BAPTIST, JEREMIAH, EVARTS, CHRISTIAN, WILLIAM, PENN, GEORGIA, WILLIAM, MCLOUGHLIN'S, GRANT, FOREMAN'S, HOLOCAUST, JAPANESE, TECUMSEH, MADISON, MONROE, JACKSON, TAHLEQUAH, B.A., M.A., PH.D., SMITHSONIAN, NATIVE (AMERICAN) STUDIES, CALIFORNIA, BERKELEY, DARTMOUTH, HANOVER, HAMPSHIRE, MINNESOTA, HASKELL, LAWRENCE, KANSAS, ALVIN, JOSEPHY, KNOPF, NANCY, SHOEMAKER, ROUTLEDGE, CHRISTIANITY, MCCLUNG, RUSSELL, THORNTON, DENVER, BOSNIA, AFRICA, EUROPEANS, MEXICO, CHIAPAS, APRIL, NOVEMBER, CHIPPEWA (total 607 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	498	722	4081	72.94%
Function:	(2385)	(42.63%)
Content:	(1696)	(30.31%)
> Anglo-Sax	(712)	(12.73%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	109	137	263	4.70%
> Anglo-Sax:	(97)	(1.73%)
1k+2k		(77.64%)
AWL Words (academic):	157	208	412	7.36%
> Anglo-Sax:	(55)	(0.98%)
Off-List Words:	?	380	839	15.00%
	764+?	1447	5595	100%
Words in text (tokens):	5595			
Different words (types):	1447			
Type-token ratio:	0.26			
Tokens per type:	3.87			
Lex density (content words/total)	0.57			

Pertaining to onlist only

Tokens:	4756
Types:	1067
Families:	764
Tokens per family:	6.23
Types per family:	1.40

Anglo-Sax Index: %
 (A-Sax tokens + functors / onlist tokens)

Greco-Lat/Fr-Cognate Index: (Inverse of above) %

Current profile	
%	Cumul.
72.94	72.94
4.70	77.64
7.36	85.00
15.00	100.00

A. AWL Tokens lists

AWL tokens [157:208:412] acquired adapt adapt adapt adapt adapt adapt
 adaptation adapted adapted adapted adapting adapting adapting affect
 annual apparent appropriate approximately area area area area areas
 areas areas aspect aspects assigned assistant assistant assistant assure
 attachment attitudes authorities authority available available bulk
 challenged challenged circumstances civil clarify commissioners commodity
 commodity communities communities communities communities community
 complex complexity conflict conflict conflict conflict constituted constitution
 consult contact contact contemporary contemporary context created
 created created created created creation cultural cultural cultural culture
 cultures debate debate debates defined defined defined defined defined
 definitive designed despite distinctive documents domain domesticated
 domesticated domesticated domesticated domesticated dominate
 dominated dominated dramatic dynamic economic economic economic
 economic economic economy economy economy edited edited element
 element emerged emerged emphasized encountered encountered
 environment environment environment environment environment
 environment environment environment environment environment
 environment environment environment environment environment
 environment environmental environments environments environments
 environments environments environments environments environments
 environments environments establish established established established
 establishing estimates ethnic ethnic evidence evident evident exploit
 exploiting expose factors factors factors federal federal federal federal
 federal final final finally finally finally financial fund generations grant grant
 guaranteed guaranteed identity identity identity identity identity identity
 identity ignored images immigrants impact impact impact impact
 implications inadequate income inherent institute institution institution
 instructor instructor integration integrity integrity interacted interpretation

interpretation invested investment involved issued journal journal
 justification liberal liberal links located maintain maintained major major
 major major margin migrants migrated migration migratory military
 motivated motives notion obviously occupied occupied occupy occurred
 parallel period period persist persist persisted persisted persistence
 perspective perspectives philosophy physical physical policy policy policy
 policy policy policy policy policy portion portions potential preceded
 previous primarily primary primary process process process processes
 promoted published pursue pursue quotes radical ranged ranges ranges
 reconstruct recreating region region region regions regions regions regions
 regions regions removal removal removal removal removal removal removal
 removal removal removal removal removal removal removal removal
 removal removal removal removal removal removal removal removal
 removal removal removal removal removal removal removal removal
 removal removal removal removal removal removal removal removal
 removal removal removals remove removed removed require required
 residence resources resources resources resources resources retain
 revolutionary revolutionary revolutionary revolutionary role role role scheme
 section secure seek seek series significance significant significant signifies
 similar similar similarities similarity site source source sources sources
 specifically stressed structures style styles styles styles submit sum
 supplement symbol technology text text text text text tradition tradition
 tradition traditional traditional traditional traditional traditional traditions
 traditions traditions transportation ultimately undergone unified unique
 utilize variations vary visualize voluntary voluntary

Sublist 1

area area area area area areas areas areas authorities authority available
 available constituted constitution context created created created created
 created creation defined defined defined defined defined defined economic
 economic economic economic economic economy economy economy
 environment environment environment environment environment
 environment environment environment environment environment
 environment environment environment environment environment
 environment environmental environments environments environments
 environments environments environments environments environments
 environments environments environments environments environments
 environments environments establish established established established
 establishing estimates evidence evident evident factors factors factors
 financial identity identity identity identity identity identity identity income
 interpretation interpretation involved issued major major major major
 occurred period period policy policy policy policy policy policy policy policy
 process process process processes recreating require required role role
 role section significance significant significant significant signifies similar similar
 similarities similarity source source sources sources specifically structures
 variations vary

Sublist 2

acquired affect appropriate aspect aspects assistant assistant assistant
commissioners communities communities communities communities
community complex complexity cultural cultural cultural culture cultures
designed distinctive element element final final finally finally finally impact
impact impact impact institute institution institution invested investment
journal journal maintain maintained potential previous primarily primary
primary ranged ranges ranges reconstruct region region regions regions
resources resources resources resources residence resources resources
resources resources resources secure seek seek site text text text text text
tradition tradition tradition traditional traditional traditional traditional
traditional traditions traditions traditions

Sublist 3

circumstances documents dominate dominated dominated emphasized
fund immigrants interacted justification links located philosophy physical
physical published removal removal removal removal removal removal
removal removal removal removal removal removal removal removal
removal removal removal removal removal removal removal removal
removal removal removal removal removal removal removal removal
removal removal removals remove removed removed scheme technology

Sublist 4

annual apparent approximately attitudes civil debate debate debates
despite domesticated domesticated domesticated domesticated
domesticated emerged emerged ethnic ethnic grant grant implications
inadequate integration obviously occupied occupied occupy parallel
promoted retain series stressed sum

Sublist 5

challenged challenged conflict conflict conflict conflict consult contact
contact expose generations images liberal liberal margin notion perspective
perspectives pursue pursue style styles styles styles symbol

Sublist 6

assigned attachment domain edited edited federal federal federal federal
federal ignored instructor instructor migrants migrated migration migratory
motivated motives preceded transportation utilize

Sublist 7

adapt adapt adapt adapt adapt adapt adaptation adapted adapted adapted
adapting adapting adapting definitive dynamic guaranteed guaranteed
quotes submit ultimately unique voluntary voluntary

Sublist 8

clarify commodity commodity contemporary contemporary dramatic exploit
exploiting radical visualize

Sublist 9

assure bulk inherent military portion portions revolutionary revolutionary
revolutionary revolutionary supplement unified

Sublist 10

encountered encountered integrity integrity persist persist persisted
persisted persistence undergone

AWL types: [157:208:412] acquired_[1] adapt_[6] adaptation_[1]
adapted_[3] adapting_[3] affect_[1] annual_[1] apparent_[1] appropriate_[1]
approximately_[1] area_[5] areas_[3] aspect_[1] aspects_[1] assigned_[1]
assistant_[3] assure_[1] attachment_[1] attitudes_[1] authorities_[1]
authority_[1] available_[2] bulk_[1] challenged_[2] circumstances_[1]
civil_[1] clarify_[1] commissioners_[1] commodity_[2] communities_[4]
community_[1] complex_[1] complexity_[1] conflict_[4] constituted_[1]
constitution_[1] consult_[1] contact_[2] contemporary_[2] context_[1]
created_[5] creation_[1] cultural_[3] culture_[1] cultures_[1] debate_[2]
debates_[1] defined_[5] definitive_[1] designed_[1] despite_[1]
distinctive_[1] documents_[1] domain_[1] domesticated_[5] dominate_[1]
dominated_[2] dramatic_[1] dynamic_[1] economic_[5] economy_[3]
edited_[2] element_[2] emerged_[2] emphasized_[1] encountered_[2]
environment_[16] environmental_[1] environments_[10] establish_[1]
established_[3] establishing_[1] estimates_[1] ethnic_[2] evidence_[1]
evident_[2] exploit_[1] exploiting_[1] expose_[1] factors_[3] federal_[5]
final_[2] finally_[3] financial_[1] fund_[1] generations_[1] grant_[2]
guaranteed_[2] identity_[7] ignored_[1] images_[1] immigrants_[1]
impact_[4] implications_[1] inadequate_[1] income_[1] inherent_[1]
institute_[1] institution_[2] instructor_[2] integration_[1] integrity_[2]
interacted_[1] interpretation_[2] invested_[1] investment_[1] involved_[1]
issued_[1] journal_[2] justification_[1] liberal_[2] links_[1] located_[1]
maintain_[1] maintained_[1] major_[4] margin_[1] migrants_[1] migrated_[1]
migration_[1] migratory_[1] military_[1] motivated_[1] motives_[1] notion_[1]
obviously_[1] occupied_[2] occupy_[1] occurred_[1] parallel_[1] period_[2]
persist_[2] persisted_[2] persistence_[1] perspective_[1] perspectives_[1]
philosophy_[1] physical_[2] policy_[8] portion_[1] portions_[1] potential_[1]
preceded_[1] previous_[1] primarily_[1] primary_[2] process_[3]
processes_[1] promoted_[1] published_[1] pursue_[2] quotes_[1] radical_[1]
ranged_[1] ranges_[2] reconstruct_[1] recreating_[1] region_[3] regions_[6]

removal_[40] removals_[1] remove_[1] removed_[2] require_[1] required_[1]
 residence_[1] resources_[5] retain_[1] revolutionary_[4] role_[3] scheme_[1]
 section_[1] secure_[1] seek_[2] series_[1] significance_[1] significant_[2]
 signifies_[1] similar_[2] similarities_[1] similarity_[1] site_[1] source_[2]
 sources_[2] specifically_[1] stressed_[1] structures_[1] style_[1] styles_[3]
 submit_[1] sum_[1] supplement_[1] symbol_[1] technology_[1] text_[5]
 tradition_[3] traditional_[5] traditions_[3] transportation_[1] ultimately_[1]
 undergone_[1] unified_[1] unique_[1] utilize_[1] variations_[1] vary_[1]
 visualize_[1] voluntary_[2]

C. AWL Families list

AWL families: [157:208:412] AWL

acquire_[1] adapt_[1] adequate_[1] affect_[1] annual_[1] apparent_[1]
 appropriate_[1] approximate_[1] area_[1] aspect_[2] assign_[1] assist_[3]
 assure_[1] attach_[1] attitude_[1] authority_[2] available_[2] bulk_[1]
 challenge_[2] circumstance_[1] civil_[1] clarify_[1] commission_[1]
 commodity_[2] community_[5] complex_[2] conflict_[4] constitute_[2]
 construct_[1] consult_[1] contact_[2] contemporary_[2] context_[1]
 create_[1] culture_[5] debate_[3] define_[5] definite_[1] design_[1]
 despite_[1] distinct_[1] document_[1] domain_[1] domestic_[5] dominate_[3]
 drama_[1] dynamic_[1] economy_[1] edit_[2] element_[2] emerge_[2]
 emphasis_[1] encounter_[2] environment_[2] establish_[5] estimate_[1]
 ethnic_[2] evident_[3] exploit_[2] expose_[1] factor_[3] federal_[5] final_[5]
 finance_[1] fund_[1] generation_[1] grant_[2] guarantee_[2] identify_[7]
 ignorant_[1] image_[1] immigrate_[1] impact_[4] implicate_[1] income_[1]
 inherent_[1] institute_[3] instruct_[2] integrate_[1] integrity_[2] interact_[1]
 interpret_[2] invest_[2] involve_[1] issue_[1] journal_[2] justify_[1] liberal_[2]
 link_[1] locate_[1] maintain_[2] major_[4] margin_[1] migrate_[4] military_[1]
 motivate_[1] motive_[1] notion_[1] obvious_[1] occupy_[3] occur_[1]
 parallel_[1] period_[2] persist_[5] perspective_[2] philosophy_[1]
 physical_[2] policy_[1] portion_[2] potential_[1] precede_[1] previous_[1]
 primary_[3] process_[4] promote_[1] publish_[1] pursue_[2] quote_[1]
 radical_[1] range_[3] region_[1] remove_[4] require_[2] reside_[1]
 resource_[5] retain_[1] revolution_[4] role_[3] scheme_[1] section_[1]
 secure_[1] seek_[2] series_[1] significant_[4] similar_[4] site_[1] source_[4]
 specific_[1] stress_[1] structure_[1] style_[4] submit_[1] sum_[1]
 supplement_[1] symbol_[1] technology_[1] text_[5] tradition_[1]
 transport_[1] ultimate_[1] undergo_[1] unify_[1] unique_[1] utilise_[1]
 vary_[2] visual_[1] voluntary_[2]

AWL Fr non-cognate families: [families 8 : tokens 55] bulk_[1] grant_[2]
 involve_[1] obvious_[1] range_[3] remove_[4] seek_[2] undergo_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	562 (49.91)	748 (51.34)	4293 (76.63)	76.63
K-2 Words :	229 (20.34)	299 (20.52)	623 (11.12)	87.75
K-3 Words :	178 (15.81)	212 (14.55)	423 (7.55)	95.30
K-4 Words :	63 (5.60)	69 (4.74)	92 (1.64)	96.94
K-5 Words :	36 (3.20)	37 (2.54)	45 (0.80)	97.74
K-6 Words :	18 (1.60)	18 (1.24)	18 (0.32)	98.06
K-7 Words :	16 (1.42)	17 (1.17)	19 (0.34)	98.40
K-8 Words :	8 (0.71)	8 (0.55)	12 (0.21)	98.61
K-9 Words :	6 (0.53)	8 (0.55)	9 (0.16)	98.77
K-10 Words :	4 (0.36)	4 (0.27)	5 (0.09)	98.86
K-11 Words :	4 (0.36)	4 (0.27)	4 (0.07)	98.93
K-12 Words :	1 (0.09)	1 (0.07)	1 (0.02)	98.95
K-13 Words :				
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :	1 (0.09)	1 (0.07)	1 (0.02)	98.97
K-24 Words :				
K-25 Words :				
Off-List:	??	13 (0.89)	17 (0.30)	99.27

Total (unrounded) 1126+? 1457 (100) 5602 (100) ≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens): 5602
 Different words (types): 1457
 Type-token ratio: 0.26
 Tokens per type: 3.84

Pertaining to onlist only

Tokens: 5585
 Types: 1444
 Families: 1126
 Tokens per Family : 4.96
 Types per Family : 1.28

Current profile (token %)	
K-1 (76.63)	76.63
K-2 (11.12)	87.75
K-3 (7.55)	95.30
K-4 (1.64)	96.94
K-5 (0.80)	97.74
K-6 (0.32)	98.06
K-7 (0.34)	98.40
K-8 (0.21)	98.61
K-9 (0.16)	98.77
K-10 (0.09)	98.86
K-11 (0.07)	98.93
K-12 (0.02)	98.95
K-23 (0.02)	98.97
OFF (0.30)	≈100%

a [91] ability [3] about [7] accept [1] act [7] action [1] actions [2] addition [1] addresses [3] afford [1] after [7] agree [1] agreeing [2] all [5] allowed [1] alone [1] along [8] already [4] also [10] although [6] among [4] an [18] and [181] animals [3] another [4] any [1] apparent [1] appears [1] are [11] area [5] areas [3] around [2] art [1] artist [4] as [36] at [14] ate [1] banks [1] based [4] be [19] bear [1] became [2] because [3] become [1] becoming [1] been [3] before [6] began [6] beginning [1] behind [3] being [1] beings [2] believed [1] below [2] between [6] beyond [3] birds [1] black [5] blood [2] board [2] boats [1] body [1] book [1] books [1] born [1] both [3] bought [1] breaks [1] broken [1] brothers [2] brought [4] build [1] buildings [1] burns [1] but [16] buy [1] by [24] called [1] came [7] camp [1] can [6] cards [1] cases [1] center [3] central [2] certain [1] change [3] changes [1] changing [2] childhood [1] children [1] chippewa [1] choice [3] choices [2] choose [1] churches [1] city [1] clothes [1] cold [1] colds [1] collection [1] collections [1] college [2] comes [1] consider [3] considered [4] continue [3] continuing [2] costly [1] could [10] country [1] courses [1] cried [1] cross [3] crossed [2] crossing [1] d [1] dance [2] day [1] days [1] deals [2] dear [1] death [1] did [2] died [1] different [3] do [4] does [2] done [1] down [3] draw [1] dream [1] dry [1] during [7] each [1] earlier [2] early [9] earth [2] easily [1] east [27] eighteenth [1] either [1] end [1] entered [2] even [1] everything [1] experience [5] experiences [1] explain [1] explained [2] explaining [1] explains [3] faced [2] fact [2] fall [1] falls [1] families [3] family [2] farm [1] farmers [1] farms [5] father [1] favorite [1] fear [2] fed [1] feel [1] felt [1] few [1] field [1] fields [3] final [2] finally [3] find [2] fires [1] first [3] five [5] flat [1] flew [1] followed [4] followers [1] food [2] for [55] force [1] forced [5] forces [3] forcibly [1] forest [2] forests [2] form [4] fought [1] found [4] free [1] freedom [2] friendly [1] from [28] front [1] full [3] further [1] game [3] gave [1] general [1] generally [7] give [2] go [2] goes [1] going [1] good [1] governing [1] government [16] great [2] green [3] grew [1] grounds [1] group [2] groups [9] grown [1] had [31] hands [1] hard [1] hardships [1] has [9] have [5] having [1] he [5] head [1] heard [1] hearts [1] held [2] help [1] her [2] here [2] hides [1] highly [1] hill [4] hills [3] hilly [2] his [9] historic [1] historical [5] history [13] hit [1] hold [1] holding [1] home [12] homeless [1] homes [7] horses [1] house [1] houses [2] how [7] however [5] human [1] humans [1] hunger [1] hunt [1] hunted [2] hunters [1] hunting [7] i [1] idea [2] ideas [1] if [2] important [2] in [167] indeed [3] interest [1] into [14] involved [1] is [23] issued [1] it [13] its [11] joining [2] joins [1] judgments [1] kind [2] known [5] land [27] lands [28] large [4] larger [3] late [4] later [2] laws [3] leader [1] leaders [1] leading [2] learned [1] leave [4] led [9] left [2] letters [2] life [7] light [1] like [1] line [1] little [2] live [6] lived [7] lives [2] living [2] local [1] long [3] low [1] m [1] made [5] madison [1] major [4] make [2] makers [1] making [3] man [2]

A. Types list

BNC-COCA-1,000 types: [fams 389 : types 522 : tokens 3837]

many [6] marked [1] market [1] may [2] mean [1] members [8] men [2] middle [1] might [2] miles [1] milk [1] minds [1] mine [1] money [5] monroe [1] monthly [1] more [2] most [5] mountain [3] mountains [6] move [10] moved [9] moving [1] much [1] must [1] nation [1] national [6] nationalism [2] nations [4] natural [5] nature [1] near [4] nearby [1] need [1] needed [1] needs [1] neighbors [1] new [26] nineteenth [2] no [2] north [12] not [13] notice [1] now [6] number [87] numbers [16] obviously [1] of [266] oil [1] on [29] one [10] ones [1] only [3] open [1] opened [2] opening [1] or [8] orders [2] other [7] others [1] our [2] ourselves [1] out [4] over [5] own [7] paid [1] painted [1] part [12] particular [2] parties [1] parts [1] passed [3] past [2] payment [2] payments [1] people [22] peoples [4] personally [1] picture [1] place [9] places [2] planted [2] plants [4] play [1] played [1] point [2] positions [1] post [1] posts [1] power [4] present [4] presented [1] press [1] pressed [1] probably [2] promise [1] promises [1] protect [2] protected [1] public [1] pushing [1] puts [1] question [1] quickly [1] rabbits [1] raise [1] raised [3] raising [2] reader [1] reading [1] reasons [3] recorded [1] red [1] relationship [3] relationships [1] renewed [2] report [1] responsibility [1] rich [1] right [1] rights [2] river [19] rivers [10] run [2] said [1] save [1] says [1] science [1] sciences [1] scientific [1] secure [1] see [5] seemingly [1] self [1] sense [5] sent [4] served [1] service [1] set [1] settled [9] settlement [1] settler [1] settlers [2] settling [2] several [1] shaped [3] shared [1] she [4] should [10] show [3] shows [1] side [1] sign [1] signed [5] simple [1] simply [2] situation [1] six [1] skins [1] sky [1] small [4] so [3] soft [2] sold [2] some [9] someone [1] song [1] south [30] space [1] spring [2] springs [1] stand [2] state [6] stations [1] step [1] still [5] stone [1] stood [2] stories [3] story [4] student [1] students [8] studies [3] study [1] subjects [1] such [8] sudden [1] sung [1] system [3] systems [1] take [1] taken [3] taught [1] tea [2] teaching [2] tears [7] ten [1] term [2] terms [1] that [57] the [502] their [79] theirs [1] them [8] there [8] these [8] they [66] thing [1] think [2] thinking [1] thirteen [1] this [8] those [8] thought [1] thousands [1] three [2] through [5] throughout [1] ties [1] time [3] tired [1] to [154] today [4] together [1] told [1] topped [1] totally [1] toward [4] towns [5] traveled [1] travels [2] treeless [1] trust [1] twenty [1] two [2] under [7] understand [2] understanding [2] unless [1] until [1] up [8] upon [1] us [2] use [5] used [4] usual [1] van [1] very [2] view [2] views [1] visiting [1] wanted [1] war [9] wars [1] was [37] washed [1] water [6] waters [2] way [4] ways [4] we [9] well [3] were [48] west [22] what [12] whatever [1] wheels [1] when [6] where [6] which [27] while [2] white [10] who [12] whole [1] whom [1] whose [2] why [1] wider [1] will [3] willing [1] wish [2] wished [1] with [31] within [2] without [2] women [4] wood [1] woods [5] word [1] words [1] work [3] worked [1]

works [1] world [6] would [14] writers [1] writings [2] written [2] yards [1] year [4] years [2] you [1] your [1]

BNC-COCA-2,000 types: [fams 230 : types 282 : tokens 628]

account [1] accounts [2] active [1] adapt [6] adaptation [1] adapted [3] adapting [3] affect [1] agents [4] april [1] arguments [2] army [1] articles [1] assistant [3] associate [1] associated [1] assure [1] attachment [1] attempt [1] attention [1] attitudes [1] available [2] balance [1] band [3] bark [1] basis [1] beans [1] bitter [1] boiled [2] branches [1] cash [1] century [5] chains [1] challenged [2] circumstances [1] cloth [1] coal [3] coast [1] common [1] communities [4] community [1] compare [1] comparisons [1] condition [1] conditions [1] connection [1] contact [2] cotton [2] coughs [2] council [1] county [2] created [5] creation [1] cultural [3] culture [1] cultures [1] currently [1] debts [1] decently [1] decisions [2] demand [1] described [3] designed [1] destroy [2] destroyed [1] detailed [1] details [1] develop [1] development [2] diet [3] directed [1] direction [2] director [2] discussion [1] disease [2] diseases [1] dramatic [1] economic [5] economy [3] edited [2] effects [1] environment [16] environmental [1] environments [10] escape [1] establish [1] established [3] establishing [1] events [2] evidence [1] exchange [5] exhaustion [1] expose [1] extend [1] extended [1] familiar [1] favored [1] fenced [1] financial [1] flooded [1] floods [1] foreign [6] fund [1] fur [1] generations [1] giant [1] grand [1] grant [2] guaranteed [2] guide [1] guiding [1] harm [1] heaven [1] identity [7] ignored [1] images [1] include [2] included [2] including [1] income [1] increasing [1] influence [2] instructor [2] interruption [1] introduced [1] introduction [1] introductions [1] iron [1] journey [2] junior [1] knowledge [2] language [2] lean [1] lesson [1] libraries [1] library [4] limited [1] located [1] log [1] loss [4] lower [6] maintain [1] maintained [1] manner [1] map [1] maps [1] march [3] meat [3] medicine [2] medicines [6] memories [1] military [1] mills [1] mission [1] missionaries [1] missionary [4] missions [2] mixed [1] modern [1] narrow [1] native [18] northern [3] november [1] objects [1] occurred [1] officials [2] oppose [1] organization [1] original [4] originally [3] path [2] peace [2] period [2] physical [2] pigs [2] piney [2] pole [3] policy [8] political [1] population [3] populations [4] possessed [1] potatoes [1] presidents [2] pressure [3] pressures [3] prevent [1] previous [1] process [3] processes [1] productive [1] productivity [1] protested [1] provided [2] quit [1] quotes [1] ranged [1] ranges [2] received [1] receives [1] recipes [1] recommend [1] recreating [1] regard [1] region [3] regions [6] regularly [1] relief [2] remain [2] remained [1] remaining [2] remarks [2] reminders [1]

removal [40] removals [1] remove [1] removed [2] repeatedly [1]
replace [1] replaced [2] representatives [1] require [1] required [1]
resisted [1] resisting [1] respecting [1] result [3] role [3] root [2] roots [1]
salt [1] satisfied [1] scenes [1] search [1] section [1] seek [2]
separated [1] september [1] series [1] shaded [1] sharp [1] similar [2]
similarities [1] similarity [1] site [1] slaves [1] social [1] societies [2]
society [6] soil [2] soldier [1] soldiers [1] southern [5] specifically [1]
spinning [1] spiritual [3] starvation [1] states [8] steam [1] stock [3]
stocked [1] storm [1] storms [1] stream [1] streams [2] stressed [1]
struggle [2] style [1] styles [3] suffered [1] supplies [1] supply [2]
technology [1] threat [3] thus [2] tips [1] trade [3] trading [1] tradition [3]
traditional [5] traditions [3] union [1] united [5] university [7]
unsuccessful [1] upper [2] valley [2] valleys [6] value [1] variations [1]
various [1] vary [1] victims [1] village [4] villages [7] violent [2] vote [1]
wandered [1] wanderers [1] wandering [1] weed [2] western [12] wing [1]

BNC-COCA-3,000 types: [fams 177 : types 203 : tokens 425]

acquired [1] acres [2] agenda [1] agriculturalists [1] agriculture [1]
alliances [1] annual [1] appropriate [1] approximately [1]
archaeological [2] aspect [1] aspects [1] assigned [1] authorities [1]
authority [1] beliefs [3] bid [1] boundary [3] campaign [1] captured [2]
cattle [4] ceremonial [1] ceremonies [3] ceremony [2] civil [1] clarify [1]
clustered [1] colonial [1] colonialism [1] colonists [1] commissioners [1]
complex [1] complexity [1] conflict [4] confronted [1] congress [7]
constituted [1] constitution [1] consult [1] contemporary [2] context [1]
convey [2] corruption [1] crisis [1] crops [6] debate [2] debates [1]
defined [5] delegates [1] dense [1] depiction [1] deposits [1]
description [2] descriptions [3] designated [1] despite [1]
disagreement [1] disrupt [1] distinctive [1] distinguished [1] divisions [2]
documents [1] dominate [1] dominated [2] eastern [17] element [2]
emerged [2] emphasized [1] encountered [2] entries [1] episodes [1]
era [1] essays [3] essential [2] estimates [1] ethnic [2] evident [2]
exploit [1] exploiting [1] exploring [1] extent [3] factors [3] faculty [1]
federal [5] fertile [1] fled [1] formation [1] grain [1] harshest [1]
headquarters [1] holy [1] hostile [1] immigrants [1] impact [4]
implications [1] inadequate [1] inhabitants [1] inspire [1] institute [1]
institution [2] integration [1] interacted [1] interpretation [2] invested [1]
investment [1] journal [2] justification [1] landscape [1] latter [1]
leased [1] liberal [2] links [1] literature [1] lobbied [1] margin [1]
migrated [1] migration [1] migratory [1] mobile [1] moral [1] motivated [1]
motives [1] museum [4] narratives [1] negotiate [1] negotiators [1]
notion [1] numerous [2] objections [1] occupied [2] occupy [1] oral [1]
origin [6] origins [1] parallel [1] persist [2] persisted [2] persistence [1]
perspective [1] perspectives [1] philosophy [1] phrase [1] portion [1]

portions [1] potential [1] preceded [1] preservation [1] preserve [1]
preserved [1] primarily [1] primary [2] professor [4] profit [1] promoted [1]
provisions [1] published [1] pursue [2] radical [1] raids [1] rebellion [1]
reconstruct [1] relatives [1] religious [3] resembled [2] residence [1]
resources [5] retain [1] revolutionary [4] scheme [1] scholars [2]
secretary [1] senses [2] significance [1] significant [2] source [2]
sources [2] sovereign [1] sovereignty [2] speculators [1] structures [1]
submit [1] succeeding [1] sum [1] superiority [1] supplement [1]
symbol [1] territorial [1] territories [2] territory [32] text [5] thorough [1]
tragedy [2] trail [7] transportation [1] treaties [8] treaty [2] tribal [17]
tribe [3] tribes [32] ultimately [1] undergone [1] unique [1] varieties [1]
vast [1] victory [1] violence [1] visualize [1] voluntary [2] wealth [1]
wealthier [1] weaving [2] yielded [1]

BNC-COCA-4,000 types: [fams 64 : types 69 : tokens 93]

affiliations [1] ancestors [1] archives [2] autonomy [1] biographies [1]
bounded [1] bulk [1] cabins [1] census [2] chapel [2] commodity [2]
comparative [1] compelled [1] compelling [1] consolidated [1]
consolidation [1] corn [6] cultivated [1] deer [4] delicate [1]
demographics [1] departure [1] dilemma [1] domain [1] dwellers [1]
dynamic [1] ecological [1] excavations [1] fort [1] furnished [1] hauls [1]
herbal [5] herbs [2] horn [1] hostage [1] inherent [1] integrity [2]
jurisdiction [1] kin [1] legislatures [1] mediators [1] mid [2] monetary [1]
necessity [2] pearl [1] portrayed [1] prey [1] reconcile [1] rigors [1]
sacred [1] savage [1] savages [1] scholarship [1] sentiments [1]
stereotypes [2] stern [1] surplus [1] terraces [1] thunder [1] timber [1]
timbers [2] tragic [1] transplanted [1] tributaries [2] unified [1] utilize [1]
virgin [1] warrant [1] warrants [1] warrior [1]

BNC-COCA-5,000 types: [fams 38 : types 38 : tokens 67]

assimilate [1] baptist [1] botanist [1] captivity [1] causal [1] clan [1]
cleansing [1] confederacy [1] conquers [1] creek [10] creeks [11]
culminated [1] definitive [1] delta [2] dependency [2] depleted [1]
emigration [1] encompasses [1] epidemic [2] expanses [1] farewell [1]
humanitarian [1] lightning [1] migrants [1] mound [2] peas [1] plains [3]
pork [1] rationale [2] roamed [1] rusty [2] signifies [1] sorrow [1]
sorrows [1] squashes [1] stewards [1] swayed [1] terrain [1] upheld [1]
upright [1]

BNC-COCA-6,000 types: [fams 18 : types 18 : tokens 18]

ambivalent [1] banished [1] covert [1] deliberation [1] humanities [1]
internment [1] lucrative [1] perish [1] pneumonia [1] proponent [1]

purifying_[1] resilience_[1] slippery_[1] squirrels_[1] stifled_[1]
transcribed_[1] uprising_[1] willow_[1]

BNC-COCA-7,000 types: [fams 17 : types 18 : tokens 22]

arid_[1] broom_[1] buffalo_[2] disparate_[1] elm_[1] excerpt_[1] foreman_[3]
lest_[1] melons_[1] motto_[1] prairie_[1] prairies_[1] reputed_[1] tonic_[1]
vanguards_[1] watershed_[1] wooded_[2] wrought_[1]

BNC-COCA-8,000 types: [fams 8 : types 8 : tokens 12]

annuity_[1] cholera_[1] dichotomy_[1] domesticated_[5] pumpkins_[1]
sill_[1] topography_[1] upcoming_[1]

BNC-COCA-9,000 types: [fams 6 : types 7 : tokens 9]

callousness_[1] cede_[1] ceding_[1] cursor_[2] dispossess_[1]
dispossessed_[1] pernicious_[1] smallpox_[1]

BNC-COCA-10,000 types: [fams 5 : types 5 : tokens 6]

buzzard_[1] duress_[2] measles_[1] stead_[1] sycamore_[1]

BNC-COCA-11,000 types: [fams 5 : types 5 : tokens]

depopulation_[1] holocaust_[1] laxative_[1] pecans_[1] yams_[1]

BNC-COCA-12,000 types: [fams 1 : types 1 : tokens]

grist_[1]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens]

pawnee_[1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams 1 : types 1 : tokens]

verdigris_[1]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams 1 : types 1 : tokens 1]

ethno_[1]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 7 : tokens 10]

canaan_[1] ceremonialism_[1] foothills_[1] forever_[1] online_[2]
railroad_[2] railroads_[2]

B. Families list

BNC-COCA-1,000 Families: [fams 389 : types 522 : tokens 3837]

a_[109] able_[3] about_[7] accept_[1] act_[10] add_[1] address_[3]
afford_[1] after_[7] agree_[3] all_[5] allow_[1] alone_[1] along_[8]
already_[4] also_[10] although_[6] among_[4] and_[181] animal_[3]
another_[4] any_[1] apparent_[1] appear_[1] area_[8] around_[2] art_[5]
as_[36] at_[14] bank_[1] base_[4] be_[145] bear_[1] because_[3]
become_[4] before_[6] begin_[7] behind_[3] believe_[1] below_[2]
between_[6] beyond_[3] bird_[1] black_[5] blood_[2] board_[2] boat_[1]
body_[1] book_[2] born_[1] both_[3] break_[2] bring_[4] brother_[2] build_[2]
burn_[1] but_[16] buy_[2] by_[24] call_[1] camp_[1] can_[6] card_[1]
case_[1] centre_[5] certain_[1] change_[6] child_[2] choice_[5] choose_[1]
church_[1] city_[1] clothes_[1] cold_[2] collect_[2] college_[2] come_[8]
consider_[7] continue_[5] cost_[1] could_[10] country_[1] course_[1]
cross_[6] cry_[1] dance_[2] day_[2] deal_[2] dear_[1] death_[1] die_[1]
different_[3] do_[9] down_[3] draw_[1] dream_[1] dry_[1] during_[7]
each_[1] early_[11] earth_[2] east_[27] easy_[1] eat_[1] eight_[1] either_[1]
end_[1] end_of_list_[1] enter_[2] even_[1] every_[1] experience_[6]
explain_[7] face_[2] fact_[2] fall_[2] family_[5] farm_[7] father_[1]
favourite_[1] fear_[2] feed_[1] feel_[2] few_[1] field_[4] fight_[1] final_[5]

find [6] fire [1] first [3] five [5] flat [1] fly [1] follow [5] food [2] for [55]
force [10] forest [4] form [4] free [3] friend [1] from [28] front [1] full [3]
further [1] game [3] general [8] give [3] go [4] good [1] govern [17]
great [2] green [3] ground [1] group [11] grow [2] hand [1] hard [2]
have [47] he [14] head [1] hear [1] heart [1] help [1] here [2] hide [1]
high [1] hill [9] history [19] hit [1] hold [4] home [20] horse [1] house [3]
how [7] however [5] human [2] hunger [1] hunt [11] i [2] idea [3] if [2]
important [2] in [167] indeed [3] interest [1] into [14] involve [1] issue [1]
it [24] join [3] judge [1] kind [2] know [5] land [55] large [7] late [6]
law [3] lead [13] learn [1] leave [4] left [2] letter [2] life [7] light [1]
like [1] line [1] little [2] live [17] local [1] long [3] low [1] major [4]
make [11] man [4] many [6] mark [1] market [1] may [2] mean [1]
member [8] middle [1] might [2] mile [1] milk [1] mind [1] money [5]
monroe [1] month [1] more [2] most [5] mountain [9] move [20] much [1]
must [1] nation [13] nature [6] near [5] need [3] neighbour [1] new [28]
nine [2] no [2] north [12] not [13] notice [1] now [6] number [103]
obvious [1] of [266] oil [1] on [29] one [11] only [3] open [4] or [8]
order [2] other [8] out [4] over [5] own [7] paint [1] part [13]
particular [2] party [1] pass [3] past [2] pay [4] people [26] person [1]
picture [1] place [11] plant [6] play [2] point [2] position [1] post [2]
power [4] present [5] press [2] probably [2] promise [2] protect [3]
public [1] push [1] put [1] question [1] quick [1] rabbit [1] raise [6]
read [2] reason [3] record [1] red [1] relate [4] report [1] responsible [1]
rich [1] right [1] rights [2] river [29] run [2] save [1] say [2] science [3]
secure [1] see [5] seem [1] self [1] sell [2] send [4] sense [5] serve [1]
service [1] set [1] settle [15] several [1] shape [3] share [1] she [6]
should [10] show [4] side [1] sign [6] simple [3] sing [1] situation [1]
six [1] skin [1] sky [1] small [4] so [3] soft [2] some [10] song [1]
south [30] space [1] spring [3] stand [4] state [6] station [1] step [1]
still [5] stone [1] story [7] student [9] study [4] subject [1] such [8]
sudden [1] system [4] take [4] tea [2] teach [3] tear [7] tell [1] ten [1]
term [3] that [65] the [502] there [8] they [154] thing [1] think [4]
thirteen [1] this [16] thousand [1] three [2] through [6] tie [1] time [3]
tire [1] to [154] today [4] together [1] top [1] total [1] toward [4] town [5]
travel [3] tree [1] trust [1] twenty [1] two [2] under [7] understand [4]
unless [1] until [1] up [8] upon [1] use [9] usual [1] van [1] very [2]
view [3] visit [1] want [1] war [10] wash [1] water [8] way [8] we [14]
well [3] west [22] what [13] wheel [1] when [6] where [6] which [27]
while [2] white [10] who [15] whole [1] why [1] wide [1] will [4] wish [3]
with [31] within [2] without [2] woman [4] wood [6] word [2] work [5]
world [6] would [14] write [5] yard [1] year [6] you [2]

BNC-COCA-2,000 Families: [fams 230 : types 282 : tokens 628]

account [3] active [1] adapt [13] affect [1] agent [4] april [1] argue [2]
army [1] article [1] assist [3] associate [2] assure [1] attach [1]
attempt [1] attention [1] attitude [1] available [2] balance [1] band [3]
bark [1] basis [1] bean [1] bitter [1] boil [2] branch [1] cash [1]
century [5] chain [1] challenge [2] circumstance [1] cloth [1] coal [3]
coast [1] common [1] community [5] compare [2] condition [2]
connect [1] contact [2] cotton [2] cough [2] council [1] county [2]
create [7] culture [5] current [1] debt [1] decent [1] decision [2]
demand [1] describe [3] design [1] destroy [3] detail [2] develop [3]
diet [3] directed [3] direction [2] discuss [1] disease [3] drama [1]
economy [8] edit [2] effect [1] environment [27] escape [1] establish [5]
event [2] evidence [1] exchange [5] exhaust [1] expose [1] extend [2]
familiar [1] favour [1] fence [1] finance [1] flood [2] foreign [6] fund [1]
fur [1] generation [1] giant [1] grand [1] grant [2] guarantee [2] guide [2]
harm [1] heaven [1] identify [7] ignore [1] image [1] include [5]
income [1] increase [1] influence [2] instruct [2] interrupt [1] introduce [3]
iron [1] journey [2] junior [1] knowledge [2] language [2] lean [1]
lesson [1] library [5] limit [1] locate [1] log [1] loss [4] lower [6]
maintain [2] manner [1] map [2] march [3] meat [3] medicine [8]
memory [1] military [1] mill [1] mission [8] mix [1] modern [1] narrow [1]
native [18] northern [3] november [1] object [1] occur [1] official [2]
oppose [1] organize [1] original [7] path [2] peace [2] period [2]
physical [2] pig [2] pine [2] pole [3] policy [8] politics [1] population [7]
possess [1] potato [1] president [2] pressure [6] prevent [1] previous [1]
process [4] product [2] protest [1] provide [2] quit [1] quote [1] range [3]
receive [2] recipe [1] recommend [1] regard [1] region [9] regular [1]
relief [2] remain [5] remark [2] remind [1] remove [44] repeat [1]
replace [3] represent [1] require [2] resist [2] respect [1] result [3]
role [3] root [3] salt [1] satisfy [1] scene [1] search [1] section [1]
seek [2] separate [1] september [1] series [1] shade [1] sharp [1]
similar [4] site [1] slave [1] social [1] society [8] soil [2] soldier [2]
southern [5] specific [1] spin [1] spirit [3] starve [1] states [8] steam [1]
stock [4] storm [2] stream [3] stress [1] struggle [2] style [4] success [1]
suffer [1] supply [3] technology [1] threat [3] thus [2] tip [1] trade [4]
tradition [11] union [1] unite [5] university [7] upper [2] valley [8]
value [1] various [1] vary [2] victim [1] village [11] violent [2] vote [1]
wander [3] weed [2] western [12] wing [1]

BNC-COCA-3,000 Families: [fams 177 : types 203 : tokens 425]

acquire [1] acre [2] adequate [1] agenda [1] agriculture [2] alliance [1]
annual [1] appropriate [1] approximate [1] archaeology [2] aspect [2]
assign [1] authority [2] belief [3] bid [1] boundary [3] campaign [1]
capture [2] cattle [4] ceremony [6] civil [1] clarify [1] cluster [1] colony [3]
commission [1] complex [2] conflict [4] confront [1] congress [7]

constitute_1] constitution_1] construct_1] consult_1] contemporary_2]
context_1] convey_2] corrupt_1] crisis_1] crop_6] debate_3] define_5]
delegate_1] dense_1] depict_1] deposit_1] description_5] designate_1]
despite_1] disagree_1] disrupt_1] distinct_1] distinguish_1] division_2]
document_1] dominate_3] eastern_17] element_2] emerge_2]
emphasise_1] encounter_2] entry_1] episode_1] era_1] essay_3]
essential_2] estimate_1] ethnic_2] evident_2] exploit_2] explore_1]
extent_3] factor_3] faculty_1] federal_5] fertile_1] flee_1] formation_1]
grain_1] harsh_1] headquarters_1] holy_1] hostile_1] immigrant_1]
impact_4] implicate_1] inhabit_1] inspire_1] institute_1] institution_2]
integrate_1] interact_1] interpret_2] invest_2] journal_2] justify_1]
landscape_1] latter_1] lease_1] liberal_2] link_1] literature_1] lobby_1]
margin_1] migrate_3] mobile_1] moral_1] motive_2] museum_4]
narrate_1] negotiate_2] notion_1] numerous_2] objected_1] occupy_3]
oral_1] origin_7] parallel_1] persist_5] perspective_2] philosophy_1]
phrase_1] portion_2] potential_1] precede_1] preserve_3] primary_3]
professor_4] profit_1] promote_1] provision_1] publish_1] pursue_2]
radical_1] raid_1] rebel_1] relatives_1] religious_3] resemble_2]
reside_1] resource_5] retain_1] revolution_4] scheme_1] scholar_2]
secretary_1] senses_2] significance_1] significant_2] source_4]
sovereign_3] speculate_1] structure_1] submit_1] succeed_1] sum_1]
superior_1] supplement_1] symbol_1] territory_35] text_5] thorough_1]
tragedy_2] trail_7] transport_1] treaty_10] tribe_52] ultimate_1]
undergo_1] unique_1] variety_1] vast_1] victory_1] violence_1]
visual_1] voluntary_2] wealth_2] weave_2] yield_1]

BNC-COCA-4,000 Families: [fams 64 : types 69 : tokens 93]

affiliate_1] ancestor_1] archive_2] autonomy_1] biography_1]
bounds_1] bulk_1] cabin_1] census_2] chapel_2] commodity_2]
comparative_1] compel_2] consolidate_2] corn_6] cultivate_1] deer_4]
delicate_1] demography_1] departure_1] dilemma_1] domain_1]
dwell_1] dynamic_1] ecological_1] excavate_1] fort_1] furnish_1]
haul_1] herb_7] horn_1] hostage_1] inherent_1] integrity_2]
jurisdiction_1] kin_1] legislature_1] mediate_1] mid_2] monetary_1]
necessity_2] pearl_1] portray_1] prey_1] reconcile_1] rigour_1]
sacred_1] savage_2] scholarship_1] sentiment_1] stereotype_2]
stern_1] surplus_1] terrace_1] thunder_1] timber_3] tragic_1]
transplant_1] tribute_2] unify_1] utilise_1] virgin_1] warrant_2]
warrior_1]

BNC-COCA-5,000 Families: [fams 38 : types 38 : tokens 67]

assimilate_1] baptist_1] botany_1] captive_1] causal_1] clan_1]
cleanse_1] confederate_1] conquer_1] creek_21] culminate_1]

definitive_1] delta_2] dependency_2] deplete_1] emigrate_1]
encompass_1] epidemic_2] expanse_1] farewell_1] humanitarian_1]
lightning_1] migrant_1] mound_2] pea_1] plains_3] pork_1] rationale_2]
roam_1] rust_2] signify_1] sorrow_2] squash_1] steward_1] sway_1]
terrain_1] uphold_1] upright_1]

BNC-COCA-6,000 Families: [fams 18 : types 18 : tokens 18]

ambivalent_1] banish_1] covert_1] deliberated_1] humanities_1]
intern_1] lucrative_1] perish_1] pneumonia_1] proponent_1] purify_1]
resilience_1] slippery_1] squirrel_1] stifle_1] transcribe_1] uprising_1]
willow_1]

BNC-COCA-7,000 Families: [fams 17 : types 18 : tokens 22]

arid_1] broom_1] buffalo_2] disparate_1] elm_1] excerpt_1] foreman_3]
lest_1] melon_1] motto_1] prairie_2] repute_1] tonic_1] vanguard_1]
watershed_1] wooded_2] wrought_1]

BNC-COCA-8,000 Families: [fams 8 : types 8 : tokens 12]

annuity_1] cholera_1] dichotomy_1] domesticate_5] pumpkin_1] sill_1]
topography_1] upcoming_1]

BNC-COCA-9,000 Families: [fams 6 : types 7 : tokens 9]

callous_1] cede_2] cursor_2] dispossess_2] pernicious_1] smallpox_1]

BNC-COCA-10,000 Families: [fams 5 : types 5 : tokens 6]

buzzard_1] duress_2] measles_1] stead_1] sycamore_1]

BNC-COCA-11,000 Families: [fams 5 : types 5 : tokens 5]

depopulate_1] holocaust_1] laxative_1] pecan_1] yam_1]

BNC-COCA-12,000 Families: [fams 1 : types 1 : tokens 1]

grist_1]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

pawnee_[1]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams 1 : types 1 : tokens 1]

verdigris_[1]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams 1 : types 1 : tokens 1]

ethno_[1]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 10 : tokens 15]

canaan_[1] ceremonialism_[1] foothills_[1] forever_[1] online_[2]
railroad_[2] railroads_[2] ♦collection_[1] ♦full_[2] ♦mcclung_[1]
♦place_[1]

7.2.3 Targets

7.2.3.1 *The Flavours of English*

7.2.3.1.1 Text only file

BEFORE YOU READ

Can you recognize any of these varieties : Canadian , Jamaican , Irish , South African , Australian and Indian English ? Revise the sounds of English with the help of the chart on page 297 .

Do the listening task on page 52 and discuss each variety before you read this text .

The Flavours of English

The USA and Canada

In the 1600s the English spoken in England and the English spoken in the New World were much the same . But as new generations grew up , the two varieties developed in different directions , especially with regard to pronunciation and vocabulary . Because American English in many ways changed less , it reflects an older , 17 and 18 century version of British English . Shakespeare and George Washington (1732 - 1799) , for example , would both have pronounced the *in* in words like farm and far and the *a* in words like after , ask and dance like an Ice / , just like American English and many British dialects today . The Americans kept the word fall , whereas the

British replaced it by autumn in the 17 century .

People from different parts of the British Isles and other countries settled in different areas of the New World . This is one of the reasons why different accents developed across the North American continent . Today there are by no means as many regional varieties of English in the USA and

Canada as there are in the British Isles , but you will no doubt have noticed

that there is a clear difference between for example a " Southern drawl " and a New York accent .

Canadian English sounds mostly like US English to us , but to Canadians there is a distinct difference . The influence from French and British English

is stronger , and this gives Canadian English some unique features .

However , young Canadians tend to be more influenced by US English and less conservative in their use of Canadian English .

Standard US English

It is rhotic (meaning *r* is clearly pronounced in most positions) .

in words and phrases like better and it is pronounced with a .

becomes in words like vast and after .

becomes in words like news and opportunity .

Standard Canadian English

Canadian pronunciation is almost identical to American pronunciation . The most famous difference is the *sound* , so the word house will sound to American ears like hoose . They tend to pronounce cot the same as caught and collar the same as caller . In Atlantic Canada , accents are more influenced by Scottish and Irish sounds .

Cockney the accent traditionally spoken by working class Londoners

A sign in pidgin English on Erakor Island , Vanuatu in the South Pacific . Can you translate the sentence into standard English ?

In Quebec people take the Metro instead of the subway , and you will hear French words such as auto route for highway , and expressions like shut a light .

Here are some slang words : clik (kilometre) , loonie (one dollar coin) ,
toonie (two dollar coin , breakwich (breakfast sandwich , Canuck
(a Canadian) , mounties (Royal Canadian Mounted Police) .

Ending sentences with " " is considered typically Canadian .

Australia and New Zealand

Compared to America , the colonization of Australia and New Zealand started much later ,
and the ties between colony and mother country have been much stronger . Many of the
convicts who came to Australia were originally from London and Ireland . Therefore ,
features of both Cockney and Irish English can be traced in Australian English today .
Australian English is in many respects similar to British English , but it also borrows from US
English . New Zealand English is different again , with many Maori loan words , but the
difference in pronunciation might only be obvious to the locals .

Received Pronunciation (' The Queen's
English ")

Pidgin a language made up of two or more languages , used as a way of communicating by
people whose first languages are different from each other .

Creole a language that is a mixture of a European language and one or more other languages
and is spoken as the first language of a people .

Australian English is non rhotic , just like Received Pronunciation , meaning that the *r* is
not pronounced in words like car and hard .

Most vowels have a different quality compared to other Englishes . For instance , the sound
often becomes *ɪ* in say , day and Australia , the first part of the diphthong being more open .
Listen carefully to an Australian saying g'day and you will notice .

Words like down and now , with a vowel next to a nasal consonant
or , are often strongly nasalized in broad Australian . This is called a " twang " .

Australians use many abbreviations , for instance : Aussie , barbie , arvo , footy , uni ,
mossie , bickie , sunnies , roo , dunny .

They also have many special slang words , such as : fair dinkum
(genuine , real) , earbashing (talking nonstop) , bloke (man) , to go bush
(leave the city) , onya (good on you well donev) a bludger (lazy person) .

The Caribbean

The varieties of English spoken here are called Caribbean English , and there is a great deal of
variation in the way it is spoken . It has been strongly influenced by the Pidgin English
spoken among the African slaves with

different native tongues . In Jamaica , for example , this has resulted in two kinds of
English : Jamaican English and Jamaican Creole . Jamaican Standard English is the language

of the government , education and newspapers . Jamaican Creole , also called Patois , is the
day to day language of the people . You might have heard this variety in reggae and ska music
, for

instance . It is not possible to understand this language without considerable effort .

Today there are more than one million people in Britain of African Caribbean descent , and
they have brought Caribbean English " back home " again . Now they are influencing British
English , especially through popular music and slang .

Jamaican English

The *r* is dropped , but not always . (It is semi rhotic .)

It has an Irish melody or intonation .

Each syllable is clearly pronounced and equally stressed , as in original .

Water is pronounced , pepper is , colour is .

Many vowels are pronounced in a different way and with a different quality .

The unvoiced sound *t* in think becomes a *ʔ* , and the voiced

sound *d* in this becomes *ʔ* .

Asia

English is widely spoken in the old colonies of the British Empire , like India

and Pakistan . Spoken Indian English naturally varies a lot across this huge

country with more than 20 officially recognized languages and more than 1 . 2 billion
inhabitants . However , there are some features that are typically Indian , for example the use
of the form : " She is knowing the answer " , and the use of non standard English tag
questions such as " You are going ,

isn't it . Indian English also has a different rhythm than British and US English . This often
makes it difficult to understand . In addition , there are major differences in pronunciation ,
vocabulary and grammar .

Indian English

is clearly pronounced after vowel sounds , like in after , sir and very .

r are pronounced without aspiration (extra breath) , like in

coolie and tiger .

l and *r* are pronounced with the tongue curled towards the back of the mouth : art .

The *θ* sound is pronounced as *ʈ* or *ʈʰ* , with the tongue curled backwards , as in : thin .

Indian English often appears to put the stress accents at other syllables than British English
. This may create a kind of " machine gun " rhythm and a clear difference in how words are
stressed and pronounced . An

example is development rather than development .

Africa

English is an official language in many African countries like South Africa, Nigeria, Zimbabwe and Sierra Leone. All of these countries have developed their own versions of English, ranging from an easily recognizable language like South African English to local pidgin and creole languages, based on English and other local African languages. English, spiced with these different local flavours, often functions as the lingua franca, at least for the educated people.

South African English uses many words from Afrikaans, like the word

lekker: South African English is lekker! The word lekker means "nice".

good, great, cool or tasty". In Nigerian Pidgin no shaking means "no cause

for alarm", and in Kenyan English do not get so lost means "don't wait so long before we see you again".

South African English (with an Afrikaans accent)

Most accents are non-rhotic, but some accents use a trilled, and sometimes a strong sound is pronounced after vowels.

The pronunciation of is different from other English varieties.

Vowel sounds: pat sounds more like pet, Africa sounds more like .

As a consequence, the vowel sounds in words like pet and pit also

change: pet sounds like it and pit like put; stars sounds like stores.

How do younger members of the Royal Family speak today?

What about BBC reporters?

The British Isles

In the United Kingdom and the Republic of Ireland there are numerous dialects and accents. Some are regional, others are social. There are not many people who actually speak Received Pronunciation, the kind of BBC English, or Queen English, that used to be a model for teaching English to foreigners. Today it is a rather posh accent, only spoken by about 2 percent of Britons in England. All cities and regions have their own accents and dialects, with Welsh, Scottish and Irish English being very distinct varieties. In South East England the most common variety is Estuary English, which shares a number of features with Cockney.

Estuary English

Words such as fast and path are pronounced with a broad : farst, parth.

is not pronounced in most words; water is pronounced .

In some words the is not pronounced, and certain words tend to end with a sound instead, faw, royal, capital

The becomes an at the end of a word: swimming swimmin.

Standard Irish English

It is rhotic (is clearly pronounced).

The Irish have a special melody in their intonation.

They have a special pronunciation of the sound .

Listen to the vowel quality in words like out, hot, wood and good.

This short overview of some of the different "Englishes" around the world makes it clear that it is the expansion of British colonial power that brought the English language to all corners of the world. In the course of the 20th century, it is the emergence of the USA as a super power that continued to strengthen the position of English as a global language. But what about the future?

International English

In many English speaking countries there are two conflicting forces. On the one hand, there is a drive towards an international standard type of English, understood by all. On the other hand, there is a wish for a unique local

variety, which preserves that country's individual identity. These local varieties can often be difficult to understand because they are heavily influenced by local languages and cultures. So in some respects English is

one language, but in other respects English really consists of many different 40 "Englishes". And where do we draw the line between a dialect and a new language? Is Ghanaian English a variety of English, a kind of dialect, or is it really a different language? How different should it be from standard International English before it is a different language? These questions are very difficult to answer, and they illustrate that the development of the English

The Circle of World English This is one way of representing the unity and diversity of the English speaking world. At the centre is placed the notion of World English, conceived as a "common core". Around it are placed the various regional or national standards, either established or becoming established. On the outside are examples of the wide range of popular Englishes which exist.

language is very complicated. Nobody can predict what the position of English will be in a hundred years time.

So in a world with a steady increase in international communication, we might need a neutral kind of standard International English. An American doing business in South Korea cannot say: "I am sure glad I did not support that boondoggle", nor can a Briton in Brazil say: "Their goals and tactics were far out of kilter". But a Norwegian in London saying "There was something muffers going on" will not be understood either. Your first, intuitive reaction in a situation where you are not understood is to repeat

your sentence more loudly and more slowly . But it is not people's hearing that is the problem in cross cultural communication . The challenge is to put yourself in the other person's shoes : how will he or she interpret what I am trying to say ? You have to realize that your way of speaking and thinking is shaped by the culture and language you have grown up with . Expressions , idioms and humour can be appreciated in one culture but often translate badly into other cultures . Therefore , you need International English ; some might call this " decaffeinated " English because it is dull as dish water . It uses words and phrases understood by all that are neither typically British nor American , nor coloured by any particular culture . The closest we get might

be international news not targeted at people in one country . But whether this kind of English will be standardized remains to be seen . So far " International English " as a kind of standard variety is more an idea than a reality .

7.2.3.1.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

You're, don't (2), I'm, didn't,

3. Hyphenated words with hyphen removed:

working-class, non-rhotic, day-to-day, African-Caribbean, semi-rhotic, o-ri-gi-nal, non-standard, machine-gun, English-speaking, cross-cultural

4. Compound words separated:

Autoroute, superpower, loanwords, dishwater

5. Words (groups of letters) removed from the text analysis:

“th” – 17th, 18th, -r, /r/, /t/, /d/, /a:/, /æ:/, /ju:/, /u:/, ou, eh, /ei/, /æi/, /n/, /r/, /wata/, /pepa/, /cola/, /O/, lo/, /d/, -ing, /k/, /p/, /p, b, t, d, k, g/, Efrica, wah-uh, royaw, capitaw,

Bullet point markers and equal signs have been removed.

6. Proper nouns:

Canadian, Jamaican, Irish, African, Australian, Indian, English, USA, Canada, England, American, British, Shakespeare, George, Washington, York, French, Canadians, Atlantic, Scottish, Irish, Zealand, Londoners, Erakor, Vanuatu, Pacific, Quebec, London, Ireland, Cockney, Maori, Caribbean, African, Jamaica, Patois, India, Pakistan, Africa, Nigeria, Zimbabwe, Sierra, Leone, Nigerian, Kenyan, Afrikaans, BBC, Britons, England, Welsh, Scottish, Estuary, Cockney, Englishes, Ghanaian, Korea, Brazil, Norwegian, Americans, European, Australians, Britain, Asia

Take note: The words outside of brackets have not been placed on the list of proper nouns.

South (African), New World, (British) Isles, New (York), US, Standard US (English), Standard (Canadian English), New (Zealand), (Erakor) Island, South (Pacific), Metro, Royal

(Canadian) Mounted Police, Received Pronunciation, (Caribbean English), Pidgin (English), (Jamaican English), (Jamaican) Creole, (Jamaican) Standard (English), (British) Empire, (Indian English), South (Africa), (Sierra Leone), South (African English), (Nigerian) Pidgin, (Kenyan English), Royal Family, (British) Isles, South East (England), (Estuary English), Standard (Irish English), International (English), (Ghanaian English), World (English), South (Korea)

Note: Text related to illustrations have been included in the text analysis.

7.2.3.1.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: The Flavours of English (13.51 kb)

Words recategorized by user as 1k items (proper nouns etc): CANADIAN, JAMAICAN, IRISH, AFRICAN, AUSTRALIAN, INDIAN, ENGLISH, USA, CANADA, ENGLAND, AMERICAN, BRITISH, SHAKESPEARE, GEORGE, WASHINGTON, YORK, FRENCH, CANADIANS, ATLANTIC, SCOTTISH, IRISH, ZEALAND, LONDONERS, ERAKOR, VANUATU, PACIFIC, QUEBEC, LONDON, IRELAND, COCKNEY, MAORI, CARIBBEAN, PIDGIN, CREOLE, AFRICAN, JAMAICA, PATOIS, INDIA, PAKISTAN, AFRICA, NIGERIA, ZIMBABWE, SIERRA, LEONE, NIGERIAN, KENYAN, AFRIKAANS, BBC, BRITONS, ENGLAND, WELSH, SCOTTISH, ESTUARY, COCKNEY, ENGLISHES, GHANAIA, KOREA, BRAZIL, NORWEGIAN, AMERICANS, EUROPEAN, AUSTRALIANS, BRITAIN, ASIA (total 228 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	301	384	1712	78.35%
Function:	(953)	(43.62%)
Content:	(759)	(34.74%)
> Anglo-Sax	(421)	(19.27%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	58	64	109	4.99%
> Anglo-Sax:	(33)	(1.51%)
1k+2k	(83.34%)
AWL Words (academic):	51	59	74	3.39%
> Anglo-Sax:	(11)	(0.50%)
Off-List Words:	?	159	290	13.27%
	410+?	666	2185	100%
Words in text (tokens):	2185			
Different words (types):	666			
Type-token ratio:	0.30			
Tokens per type:	3.28			

Lex density (content words/total) 0.56

Pertaining to onlist only

Tokens: 1895
 Types: 507
 Families: 410
 Tokens per family: 4.62
 Types per family: 1.24
 Anglo-Sax Index: %
(A-Sax tokens + functors / onlist tokens)
 Greco-Lat/Fr-Cognate Index: (Inverse of above) %

Current profile	
%	Cumul.
78.35	78.35
4.99	83.34
3.39	86.73
13.27	100.00

A. AWL Tokens lists

AWL [51:59:74] appreciated areas challenge chart communicating communication communication conceived conflicting consequence considerable consists core create cultural culture culture culture cultures cultures distinct distinct diversity emergence established established expansion features features features features functions generations global goals identical identity illustrate individual instance instance instance interpret major neutral notion obvious percent predict range ranging reaction regional regional regional regions revise route similar stress stressed stressed targeted task text traced traditionally unique unique variation varies version versions whereas

Sublist 1

areas consists create established established functions identity individual interpret major percent similar variation varies

Sublist 2

consequence cultural culture culture culture cultures cultures distinct distinct features features features features range ranging regional regional regional regions text traditionally

Sublist 3

considerable core illustrate instance instance instance reaction task

Sublist 4

communicating communication communication emergence goals obvious predict stress stressed stressed

Sublist 5

challenge conflicting expansion generations notion targeted version versions whereas

Sublist 6

diversity neutral traced

Sublist 7

global identical unique unique

Sublist 8

appreciated chart revise

Sublist 9

route

Sublist 10

conceived

B. AWL Types list

AWL types: [51:59:74] appreciated_[1] areas_[1] challenge_[1] chart_[1] communicating_[1] communication_[2] conceived_[1] conflicting_[1] consequence_[1] considerable_[1] consists_[1] core_[1] create_[1] cultural_[1] culture_[3] cultures_[2] distinct_[2] diversity_[1] emergence_[1] established_[2] expansion_[1] features_[4] functions_[1] generations_[1] global_[1] goals_[1] identical_[1] identity_[1] illustrate_[1] individual_[1] instance_[3] interpret_[1] major_[1] neutral_[1] notion_[1] obvious_[1] percent_[1] predict_[1] range_[1] ranging_[1] reaction_[1] regional_[3] regions_[1] revise_[1] route_[1] similar_[1] stress_[1] stressed_[2] targeted_[1] task_[1] text_[1] traced_[1] traditionally_[1] unique_[2] variation_[1] varies_[1] version_[1] versions_[1] whereas_[1]

C. AWL Families list

appreciate_[1] area_[1] challenge_[1] chart_[1] communicate_[3] conceive_[1] conflict_[1] consequent_[1] considerable_[1] consist_[1] core_[1] create_[1] culture_[6] distinct_[2] diverse_[1] emerge_[1] establish_[2] expand_[1] feature_[4] function_[1] generation_[1] globe_[1] goal_[1] identical_[1] identify_[1] illustrate_[1] individual_[1] instance_[3] interpret_[1] major_[1] neutral_[1] notion_[1] obvious_[1] percent_[1] predict_[1] range_[2] react_[1] region_[4] revise_[1] route_[1] similar_[1]

stress_[3] target_[1] task_[1] text_[1] trace_[1] tradition_[1] unique_[2]
vary_[2] version_[2] whereas_[1]

AWL Fr non-cognate families: [families 7: tokens 11] core_[1] feature_[4]
goal_[1] obvious_[1] range_[2] target_[1] whereas_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: The Flavours of English (13,878 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): canadian, jamaican, irish, african, australian, indian, english, usa, canada, england, american, british, shakespeare, george, washington, york, french, canadians, atlantic, scottish, irish, zealand, londoners, erakor, vanuatu, pacific, quebec, london, ireland, cockney, maori, caribbean, pidgin, creole, african, jamaica, patois, india, pakistan, africa, nigeria, zimbabwe, sierra, leone, nigerian, kenyan, afrikaans, bbc, britons, england, welsh, scottish, estuary, cockney, englishes, ghanaiian, korea, brazil, norwegian, americans, european, australians, britain, asia end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	195 (62.70)	223 (63.17)	896 (77.17)	77.17
K-2 Words :	58 (18.65)	63 (17.85)	152 (13.09)	90.26
K-3 Words :	41 (13.18)	45 (12.75)	83 (7.15)	97.41
K-4 Words :	5 (1.61)	5 (1.42)	5 (0.43)	97.84
K-5 Words :	5 (1.61)	5 (1.42)	5 (0.43)	98.27
K-6 Words :	1 (0.32)	1 (0.28)	1 (0.09)	98.36
K-7 Words :	2 (0.64)	2 (0.57)	2 (0.17)	98.53
K-8 Words :				
K-9 Words :				
K-10 Words :				
K-11 Words :	1 (0.32)	1 (0.28)	1 (0.09)	98.62
K-12 Words :				
K-13 Words :				

K-14 Words :				
K-15 Words :	2 (0.64)	2 (0.57)	6 (0.52)	99.14
K-16 Words :				
K-17 Words :	1 (0.32)	1 (0.28)	1 (0.09)	99.23
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	3 (0.85)	7 (0.60)	99.83
Total (unrounded)	311+?	353 (100)	1161 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1161
Different words (types):	353
Type-token ratio:	0.30
Tokens per type:	3.29

Pertaining to onlist only

Tokens:	1154
Types:	350
Families:	311
Tokens per Family :	3.71
Types per Family :	1.13

Current profile (token %)	
K-1 (77.17)	77.17
K-2 (13.09)	90.26
K-3 (7.15)	97.41

K-4 (0.43)	97.84
K-5 (0.43)	98.27
K-6 (0.09)	98.36
K-7 (0.17)	98.53
K-11 (0.09)	98.62
K-15 (0.52)	99.14
K-17 (0.09)	99.23
OFF (0.60)	≈100%

A. Types list

Highlighted words:

Green = words in the glossary

Gray= AWL vocabulary in the text

Pink= AWL vocabulary in the glossary

Blue= parts of collocations in the glossary

BNC-COCA-1,000 types: [fams 260 : types 307 : tokens 1540]

a [60] about [3] across [2] actually [1] addition [1] after [5] again [3]
 all [5] almost [1] also [5] always [1] am [2] among [1] an [12] and [82]
 answer [2] any [2] appears [1] are [30] areas [1] around [2] art [1]
 as [21] asia [1] ask [1] at [5] autumn [1] back [2] backwards [1]
 badly [1] based [1] be [12] because [3] becomes [6] becoming [1]
 been [2] before [4] being [2] better [1] between [3] billion [1] both [2]
 breakfast [1] breath [1] brought [2] bush [1] business [1] but [13] by [13]
 call [1] called [3] caller [1] came [1] can [7] cannot [1] car [1]
 carefully [1] caught [1] cause [1] centre [1] certain [1] change [1]
 changed [1] cities [1] city [1] class [1] clear [3] clearly [4] closest [1]
 colour [1] coloured [1] considered [1] continued [1] cool [1] corners [1]
 countries [4] country [4] course [1] cross [1] dance [1] day [4] deal [1]
 did [1] difference [5] differences [1] different [18] difficult [3] do [4]
 doing [1] don [1] doubt [1] down [1] draw [1] drive [1] dropped [1]
 each [3] ears [1] easily [1] east [1] educated [1] education [1] either [2]
 end [2] ending [1] especially [2] expressions [2] extra [1] fair [1] fall [1]
 family [1] far [3] farm [1] fast [1] first [4] for [12] forces [1] form [1]
 from [9] get [2] gives [1] glad [1] go [1] going [2] good [3]
 government [1] great [2] grew [1] grown [1] gun [1] hand [2] hard [1]
 has [4] have [13] he [1] hear [1] heard [1] hearing [1] heavily [1]
 help [1] here [2] home [1] hot [1] house [1] how [4] however [2]
 huge [1] hundred [1] i [3] ice [1] idea [1] in [68] instead [2] into [2]
 is [63] island [1] it [25] just [2] kept [1] kind [6] kinds [1] kingdom [1]
 knowing [1] later [1] lazy [1] least [1] leave [1] less [2] light [1] like [26]
 line [1] listen [2] listening [1] local [6] locals [1] long [1] lost [1] lot [1]

loudly [1] machine [1] made [1] major [1] makes [2] man [1] many [14]
 may [1] meaning [2] means [4] members [1] might [5] million [1]
 more [13] most [6] mostly [1] mother [1] mouth [1] much [3] music [2]
 national [1] naturally [1] need [2] new [8] news [2] next [1] nice [1]
 no [4] nobody [1] nonstop [1] north [1] not [14] notice [1] noticed [1]
 now [2] number [13] numbers [1] obvious [1] of [64] often [7] old [1]
 older [1] on [9] one [9] only [2] open [1] or [11] other [11] others [1]
 out [2] outside [1] own [2] page [2] part [1] particular [1] parts [1]
 people [10] person [2] placed [2] police [1] position [2] positions [1]
 possible [1] power [2] problem [1] put [3] queen [2] questions [2]
 rather [2] read [2] real [1] reality [1] realize [1] really [2] reasons [1]
 reporters [1] same [3] say [4] saying [2] see [1] seen [1] settled [1]
 shaking [1] shaped [1] shares [1] she [2] shoes [1] short [1] should [1]
 shut [1] sign [1] sir [1] situation [1] slowly [1] so [6] some [9]
 something [1] sometimes [1] sound [9] sounds [10] south [9] speak [2]
 speaking [3] special [3] spoken [10] stars [1] started [1] stores [1]
 strong [1] stronger [2] strongly [2] such [4] support [1] sure [1]
 swimming [1] take [1] talking [1] tasty [1] teaching [1] tend [3] than [7]
 that [14] the [119] their [5] there [14] these [5] they [7] think [1]
 thinking [1] this [15] through [1] ties [1] time [1] to [33] today [6]
 towards [2] trying [1] two [5] type [1] understand [3] understood [4]
 up [3] us [6] use [5] used [2] uses [2] very [4] voiced [1] wait [1]
 was [1] water [3] way [5] ways [1] we [4] well [1] were [3] what [4]
 where [2] whether [1] which [3] who [2] whose [1] why [1] wide [1]
 widely [1] will [8] wish [1] with [19] without [2] wood [1] word [5]
 words [20] working [1] world [8] would [1] years [1] you [15] young [1]
 younger [1] your [3] yourself [1]

BNC-COCA-2,000 types: [fams 93 : types 109 : tokens 232]

accent [4] accents [7] alarm [1] appreciated [1] borrows [1] [redacted] [2]
 capital [1] century [3] challenge [1] circle [1] common [2] compared [2]
 complicated [1] create [1] cultural [1] culture [3] cultures [2] curled [2]
 developed [3] development [3] directions [1] discuss [1] dish [1]
 dollar [2] effort [1] empire [1] equally [1] established [2] example [5]
 examples [1] exist [1] famous [1] features [4] foreigners [1] future [1]
 generations [1] goals [1] identity [1] illustrate [1] increase [1]
 individual [1] influence [1] influenced [4] influencing [1] instance [3]
 kilometre [1] language [17] languages [7] loan [1] model [1] mounted [1]
 native [1] neither [1] newspapers [1] non [3] nor [3] official [1]
 officially [1] opportunity [1] original [1] originally [1] pat [1] path [1]
 percent [1] pet [3] popular [2] pronounce [1] pronounced [18]
 pronunciation [10] quality [3] range [1] ranging [1] reaction [1]
 received [3] recognizable [1] recognize [1] recognized [1] regard [1]
 regional [3] regions [1] remains [1] repeat [1] replaced [1]

representing_1] respects_3] resulted_1] royal_3] sandwich_1]
sentence_2] sentences_1] similar_1] slaves_1] social_1] southern_1]
standard_10] standardized_1] standards_1] steady_1] strengthen_1]
stress_1] stressed_2] super_1] therefore_2] thin_1] tongue_2]
tongues_1] traced_1] traditionally_1] typically_3] united_1] variation_1]
varies_1] various_1] version_1] versions_1] whereas_1]

BNC-COCA-3,000 types: [fams 49 : types 54 : tokens 81]

chart_1] coin_2] colonial_1] colonies_1] colonization_1] colony_1]
communicating_1] communication_2] conceived_1] conflicting_1]
consequence_1] conservative_1] considerable_1] consists_1]
continent_1] convicts_1] core_1] distinct_2] diversity_1] emergence_1]
expansion_1] flavours_2] functions_1] genuine_1] global_1] highway_1]
humour_1] inhabitants_1] international_8] interpret_1] mixture_1]
neutral_1] notion_1] numerous_1] pepper_1] phrases_2] pit_2]
predict_1] preserves_1] reflects_1] republic_1] revise_1] rhythm_2]
route_1] tactics_1] targeted_1] task_1] text_1] translate_2] unique_2]
unity_1] varieties_7] variety_6] vast_1]

BNC-COCA-4,000 types: [fams 10 : types 10 : tokens 11]

aspiration_1] auto_1] click_1] collar_1] dull_1] identical_1] melody_2]
spiced_1] tea_1] tiger_1]

BNC-COCA-5,000 types: [fams 4 : types 4 : tokens 5]

descent_1] grammar_1] overview_1] vocabulary_2]

BNC-COCA-6,000 types: [fams 4 : types 5 : tokens 11]

dialect_2] dialects_3] intuitive_1] isles_3] syllable_1] syllables_1]

BNC-COCA-7,000 types: [fams 11 : types 12 : tokens 24]

abbreviations_1] bloke_1] cot_1] creole_4] drawl_1] estuary_2]
idioms_1] intonation_2] nasal_1] nasalized_1] subway_1] vowel_5]
vowels_3]

BNC-COCA-8,000 types: [fams 4 : types 4 : tokens 6]

consonant_1] metro_1] posh_1] slang_3]

BNC-COCA-9,000 types: [fams 3 : types 3 : tokens 3]

reggae_1] roo_1] semi_1]

BNC-COCA-10,000 types: [fams 2 : types 2 : tokens 2]

loonie_1] [redacted]_1]

BNC-COCA-11,000 types: [fams 2 : types 2 : tokens 4]

cockney_3] twang_1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams 3 : types 3 : tokens]

coolie_1] decaffeinated_1] patois_1]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 5 : types 5 : tokens]

diphthong_1] like_1] lingua_1] pidgin_5] uni_1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams 3 : types 3 : tokens 3]

boondoggle_1] footy_1] mounties_1]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams 1 : types 1 : tokens 1]

dinkum_1]

BNC-COCA-25,000 types: [fams 1 : types 1 : tokens 1]

arvo_1]

OFFLIST: [?: types 26 : tokens 32]

aussie [1] barbie [1] bickie [1] bludger [1] breakwich [1] canuck [1]
donev [1] dunny [1] earbashing [1] farst [1] faw [1] franca [1] hoose [1]
lekker [3] mossie [1] muffens [1] numero [1] onya [1] parth [1] **molio** [5]
ska [1] sunnies [1] swimmin [1] toonie [1] unvoiced [1]

B. Families list

BNC-COCA-1,000 Families: [fams 260 : types 307 : tokens 1540]

a [72] about [3] across [2] actual [1] add [1] after [5] again [3] all [5]
almost [1] also [5] always [1] among [1] and [82] answer [2] any [2]
appear [1] area [1] around [2] art [1] as [21] ask [1] at [5] autumn [1]
back [3] bad [1] base [1] be [115] because [3] become [7] before [4]
better [1] between [3] billion [1] both [2] breakfast [1] breath [1] bring [2]
bush [1] business [1] but [13] by [13] call [5] can [8] car [1] care [1]
catch [1] cause [1] centre [1] certain [1] change [2] city [2] class [1]
clear [7] close [1] colour [2] come [1] consider [1] continue [1] cool [1]
corner [1] country [8] course [1] cross [1] dance [1] day [4] deal [1]
difference [6] different [18] difficult [3] do [7] doubt [1] down [1] draw [1]
drive [1] drop [1] each [3] ear [1] east [1] easy [1] educate [2] either [2]
end [3] end_of_list [1] especially [2] express [2] extra [1] fair [1] fall [1]
family [1] far [3] farm [1] fast [1] first [4] for [12] force [1] form [1]
from [9] get [2] give [1] glad [1] go [3] good [3] govern [1] great [2]
grow [2] gun [1] hand [2] hard [1] have [17] he [1] hear [3] heavy [1]
help [1] here [2] home [1] hot [1] house [1] how [4] however [2]
huge [1] hundred [1] i [3] ice [1] idea [1] in [68] instead [2] into [2]
island [1] it [25] just [2] keep [1] kind [7] king [1] know [1] late [1]
lazy [1] least [1] leave [1] less [2] light [1] like [26] line [1] listen [3]
local [7] long [1] lose [1] lot [1] loud [1] machine [1] major [1] make [3]
man [1] many [14] may [1] mean [6] member [1] might [5] million [1]
more [13] most [7] mother [1] mouth [1] much [3] music [2] nation [1]
nature [1] need [2] new [8] news [2] next [1] nice [1] no [4] nobody [1]
north [1] not [14] notice [2] now [2] number [14] obvious [1] of [64]
often [7] old [2] on [9] one [9] only [2] open [1] or [11] other [12] out [3]
own [2] page [2] part [2] particular [1] people [10] person [2] place [2]
police [1] position [3] possible [1] power [2] problem [1] put [3] queen [2]
question [2] rather [2] read [2] real [2] realise [1] really [2] reason [1]
report [1] same [3] say [6] see [2] settle [1] shake [1] shape [1]
share [1] she [2] shoe [1] short [1] should [1] shut [1] sign [1] sir [1]
situation [1] slow [1] so [6] some [11] sound [19] south [9] speak [15]
special [3] star [1] start [1] stop [1] store [1] strong [5] such [4]
support [1] sure [1] swim [1] take [1] talk [1] taste [1] teach [1] tend [3]
than [7] that [14] the [119] there [14] they [12] think [2] this [20]
through [1] tie [1] time [1] to [33] today [6] toward [2] try [1] two [5]

type [1] understand [7] up [3] use [9] very [4] voice [1] wait [1] water [3]
way [6] we [10] well [1] what [4] where [2] whether [1] which [3] who [3]
why [1] wide [2] will [8] wish [1] with [19] without [2] wood [1] word [25]
work [1] world [8] would [1] year [1] you [19] young [2]

BNC-COCA-2,000 Families: [fams 93 : types 109 : tokens 232]

accent [11] alarm [1] appreciate [1] borrow [1] broad [2] capital [1]
century [3] challenge [1] circle [1] common [2] compare [2]
complicate [1] create [1] culture [6] curl [2] develop [6] direction [1]
discuss [1] dish [1] dollar [2] effort [1] empire [1] equal [1] establish [2]
example [6] exist [1] famous [1] feature [4] foreign [1] future [1]
generation [1] goal [1] identify [1] illustrate [1] increase [1] individual [1]
influence [6] instance [3] kilometre [1] language [24] loan [1] model [1]
mount [1] **native** [1] neither [1] newspaper [1] non [3] nor [3] official [2]
opportunity [1] original [2] pat [1] path [1] percent [1] pet [3] popular [2]
pronounce [29] quality [3] range [2] react [1] receive [3] recognize [3]
regard [1] region [4] remain [1] repeat [1] replace [1] represent [1]
respect [3] result [1] royal [3] sandwich [1] sentence [3] similar [1]
slave [1] social [1] southern [1] standard [12] steady [1] strength [1]
stress [3] super [1] therefore [2] thin [1] **tongue** [3] trace [1] tradition [1]
typical [3] unite [1] various [1] vary [2] version [2] whereas [1]

BNC-COCA-3,000 Families: [fams 49 : types 54 : tokens 81]

chart [1] coin [2] colony [4] communicate [3] conceive [1] conflict [1]
consequence [1] conservative [1] considerable [1] consist [1]
continent [1] convict [1] core [1] distinct [2] diverse [1] emerge [1]
expansion [1] flavour [2] function [1] genuine [1] global [1] highway [1]
humour [1] inhabit [1] international [8] interpret [1] mixture [1] neutral [1]
notion [1] numerous [1] pepper [1] phrase [2] pit [2] predict [1]
preserve [1] reflect [1] republic [1] revise [1] rhythm [2] route [1]
tactic [1] target [1] task [1] text [1] translate [2] unique [2] unity [1]
variety [13] vast [1]

BNC-COCA-4,000 Families: [fams 10 : types 10 : tokens 11]

aspire [1] automobile [1] click [1] collar [1] dull [1] identical [1]
melody [2] spice [1] tag [1] tiger [1]

BNC-COCA-5,000 Families: [fams 4 : types 4 : tokens 5]

descent [1] grammar [1] overview [1] vocabulary [2]

BNC-COCA-6,000 Families: [fams 4 : types 5 : tokens 11]

dialect_5] intuitive_1] isle_3] syllable_2]

BNC-COCA-7,000 Families: [fams 11 : types 12 : tokens 24]

abbreviate_1] bloke_1] cot_1] creole_4] drawl_1] estuary_2] idiom_1]
intone_2] nasal_2] subway_1] vowel_8]

BNC-COCA-8,000 Families: [fams 4 : types 4 : tokens 6]

consonant_1] metro_1] posh_1] slang_3]

BNC-COCA-9,000 Families: [fams 3 : types 3 : tokens 3]

kangaroo_1] reggae_1] semi_1]

BNC-COCA-10,000 Families: [fams 2 : types 2 : tokens 2]

loony_1] trill_1]

BNC-COCA-11,000 Families: [fams 2 : types 2 : tokens 4]

cockney_3] twang_1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams 3 : types 3 : tokens 3]

caffein_1] coolie_1] patois_1]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 5 : types 5 : tokens 9]

diphthong_1] kilter_1] lingua_1] pidgin_5] uni_1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams 3 : types 3 : tokens 3]

boondoggle_1] footy_1] mountie_1]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams 1 : types 1 : tokens 1]

dinkum_1]

BNC-COCA-25,000 Families: [fams 1 : types 1 : tokens 1]

arvo_1]

OFFLIST: [?: types 26 : tokens 32]

aussie_1] barbie_1] bickie_1] bludger_1] breakwich_1] canuck_1]
donev_1] dunny_1] earbashing_1] farst_1] faw_1] franca_1] hoose_1]
lekker_3] mossie_1] muffens_1] numero_1] onya_1] parth_1] rhotic_5]
ska_1] sunnies_1] swimmin_1] toonie_1] unvoiced_1]

7.2.3.2 *The Power of English* Part 1

7.2.3.2.1 Text only file

BEFORE YOU READ

What does it mean that English is a universal or global language ?

Is English the only world language ?

CENTRAL SUBCOMMITTEE STEERING GROUP FOR THE FACILITATION OF
BREVITY AND CLARITY OF LANGUAGE USAGE IN THE USER / PROVIDER
COMMUNICATION INTERFACE

(formerly the plain English group)

The Power of English , Part 1

Introduction

English has become a global or universal language because of its power , or more precisely ,
because it was the language of the people with power . As a result , it has become a powerful
language , a language that gives strength and opportunity to the people who master it . So the
power of English is twofold :

The language itself has influence now that it has spread all over the world and has become
the common international means of communication .

The language empowers the people who can use it, as better skills and education can open up job prospects and increase their standard of living.

English is defined as a universal or global language because it is spoken all across the globe and is used for communication world wide by people with different native tongues. In other words, it is a matter of numbers and of geographical distribution. In addition, we have to look at what the language is used for: which international organizations and industries use the language, and how important is the language for communication across borders? Who are the people that use the language, and what do they use it for?

Numbers

As of today, it is only English that fulfils all the criteria of a global language: it is the most widely spoken language in the world. But this is only true as long as we count both native speakers and second and third language speakers of English. It is estimated that there are between 335 and 430 million native speakers of English, and presumably more than 500 million who use English as a second language. Furthermore, there are all the

people who can speak some English: yes, no, thank you, bye bye. Estimates range from 500 to 1000 million, for what does it really mean to speak English? It is guesstimated that at least one in five of the world's population speaks English at a good level of competence and that non native speakers now outnumber native speakers by a ratio of 3 to 1. All in all, there are definitely more than a billion people who speak English world wide. If you include all those who speak some English or try to speak a little English,

who knows what the number is? And besides, what is English? Is Jamaican Creole English or not? And what about Nigerian Pidgin?

There are also other languages with large numbers of speakers in different countries, but making statistics is tricky business. The most important and most difficult question is: what exactly is one language? Are Norwegian, Swedish and Danish three languages or three variants of one language, since they are really mutually intelligible? When does a variety of a language become a new language? And what about the difference between spoken and written language? If you take into account the uncertainty of the definition of a language, here are estimates of the nine

most widely spoken (not written) languages in the world by number of native speakers:

To what extent are Chinese and Spanish global languages?

How many countries do you know that fit into the different categories?

Geographical distribution

English is the primary language (spoken natively by the majority of the population) in only six major countries: the USA, Canada, the UK, Ireland, Australia and New Zealand. In addition, English is the primary language in many smaller countries and territories, especially in the Caribbean. There are also many countries where English is an official language, yet not the most widely spoken language, for instance India and South Africa. All in all, there are some 95 sovereign states and non sovereign entities (such as Hong Kong, Bermuda and Puerto Rico) where English is an official or the

dominant language, either by law or in practice.

There are three types of countries:

English is the native and first language of the majority of the population.

English is not the native tongue, but is important for historical reasons and plays a part in the nation's institutions, either as an official language or otherwise. It is a second language.

English plays no historical or governmental role, but it is nevertheless widely used as a foreign language or lingua franca, and it is taught as a foreign language in schools.

So English is represented in every continent and in the three major regions, and this geographical spread justifies the use of the label "global language".

British English

How much is 10% of the Indian population? How does this number compare to the number of native speakers in the UK and the USA? Could this have an influence on the future of the English language?

Do you know of any more areas where English is the natural language of communication? Do you know of any areas where English is NOT used for international communication?

Lingua Franca

English is also a global language because it is a lingua franca, a common language that enables people from diverse backgrounds and ethnicities to communicate when they have different native tongues.

Take India as an example. India has hundreds of languages but no national language. There are two official languages: Hindi and English. English is the most widely used lingua franca and the language of education, especially higher education. Mastering English is the ticket out of poverty for many Indians. It is also extensively used in the media, the

central government and commerce. English language literature, film television, music and theatre are also popular. Yet 70-80 percent of the population do not speak English, and only an estimated 10 percent really qualify as speakers of English as an additional language.

English has also become the lingua franca of the world and of international communication in many different fields, among them:

one of the official languages of the United Nations and most other international organizations

the official language of international air traffic control and sea faring communications

the most commonly used language of international diplomacy (together with French)

the dominant language in science, technology and academics

the preferred language of international business and trade, industry

and finance

the dominant language of the international press, media, TV and radio

the language of computing and the internet (but less so than previously)

the language of international travel and tourism

the language of the international entertainment and film industry

the dominant language of popular music

the language of international sports events and competitions

The use of English in so many fields contributes to a further strengthening of the position of English , giving the language even more power and giving the people who master the language more power . It also promotes the continued growth and development of English as a global language .

7.2.3.2.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

non-native, non-sovereign, English-language,

4. Compound words separated:

worldwide (2), seafaring

5. Words (groups of letters) removed from the text analysis:

Bullet points have also been removed from the text.

6. Proper nouns:

English, Jamaican, Nigerian, Norwegian, Swedish, Danish, USA, Canada, UK, Ireland, Australia, Zealand, Caribbean, India, Africa, Hong, Kong, Bermuda, Puerto, Rico, Hindi, French, TV, Spanish, Indian, Indians, Chinese, British

Take note: The words outside of brackets have not been placed on the list of proper nouns.

(Jamaican) Creole, (Nigerian) Pidgin, New (Zealand), South (Africa)

Note: Text related to illustrations have been included in the text analysis.

7.2.3.2.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Untitled (7.15 kb)

Words recategorized by user as 1k items (proper nouns etc): ENGLISH, JAMAICAN, NIGERIAN, NORWEGIAN, SWEDISH, DANISH, USA, CANADA, UK, IRELAND, AUSTRALIA, ZEALAND, CARIBBEAN, INDIA, AFRICA, HONG, KONG, BERMUDA, PUERTO, RICO, HINDI, FRENCH, TV, SPANISH, INDIAN, INDIANS, CHINESE, BRITISH (total 73 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	186	220	969	83.46%
Function:	(534)	(45.99%)
Content:	(435)	(37.47%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(189)	(16.28%)
K2 Words (1001-2000):	24	26	44	3.79%
> Anglo-Sax:	(7)	(0.60%)
1k+2k	(87.25%)
AWL Words (academic):	44	50	75	6.46%
> Anglo-Sax:	(3)	(0.26%)
Off-List Words:	<u>2</u>	<u>56</u>	<u>73</u>	<u>6.29%</u>
	254+?	352	1161	100%

Words in text (tokens):	1161
Different words (types):	352
Type-token ratio:	0.30
Tokens per type:	3.30
Lex density (content words/total)	0.54

Pertaining to onlist only

Tokens:	1088
Types:	296
Families:	254
Tokens per family:	4.28
Types per family:	1.17
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
83.46	83.46
3.79	87.25

6.46	93.71
6.29	100.00

A. AWL Tokens lists

AWL [44:50:75] academics areas areas brevity categories clarity communicate communication communication communication communication communication computing contributes criteria defined definitely definition distribution distribution diverse dominant dominant dominant enables entities estimated estimated estimates estimates facilitation finance furthermore global global global global global global globe instance institutions job justifies label major major majority majority media media mutually nevertheless percent percent precisely presumably previously primary primary promotes prospects range ratio regions role statistics technology variants

Sublist 1

areas areas defined definition distribution distribution estimated estimated estimates estimates finance major major majority majority percent percent role variants

Sublist 2

categories computing institutions previously primary primary range regions

Sublist 3

contributes criteria dominant dominant dominant dominant instance justifies technology

Sublist 4

communicate communication communication communication communication communication communication communication communications job label promotes statistics

Sublist 5

academics enables entities facilitation precisely ratio

Sublist 6

brevity diverse furthermore nevertheless presumably

Sublist 7

definitely global global global global global global globe media media

Sublist 8

clarity prospects

Sublist 9

mutually

B. AWL Types list

AWL types: [44:50:75] academics_[1] areas_[2] brevity_[1] categories_[1] clarity_[1] communicate_[1] communication_[7] communications_[1] computing_[1] contributes_[1] criteria_[1] defined_[1] definitely_[1] definition_[1] distribution_[2] diverse_[1] dominant_[4] enables_[1] entities_[1] estimated_[2] estimates_[2] facilitation_[1] finance_[1] furthermore_[1] global_[8] globe_[1] instance_[1] institutions_[1] job_[1] justifies_[1] label_[1] major_[2] majority_[2] media_[2] mutually_[1] nevertheless_[1] percent_[2] precisely_[1] presumably_[1] previously_[1] primary_[2] promotes_[1] prospects_[1] range_[1] ratio_[1] regions_[1] role_[1] statistics_[1] technology_[1] variants_[1]

C. AWL Families list

AWL families: [44:50:75] academy_[1] area_[2] brief_[1] category_[1] clarify_[1] communicate_[9] compute_[1] contribute_[1] criteria_[1] define_[2] definite_[1] distribute_[2] diverse_[1] dominate_[4] enable_[1] entity_[1] estimate_[4] facilitate_[1] finance_[1] furthermore_[1] globe_[9] instance_[1] institute_[1] job_[1] justify_[1] label_[1] major_[4] media_[2] mutual_[1] nevertheless_[1] percent_[2] precise_[1] presume_[1] previous_[1] primary_[2] promote_[1] prospect_[1] range_[1] ratio_[1] region_[1] role_[1] statistic_[1] technology_[1] vary_[1]

AWL Fr non-cognate families: [families 3 : tokens 3] furthermore_[1] nevertheless_[1] range_[1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, propers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: The Power of English Part 1 (7,329 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): english, jamaican, nigerian, norwegian, swedish, danish, usa, canada, uk, ireland, australia, zealand, caribbean, india, africa, hong, kong, bermuda, puerto rico, hindi, french, tv, spanish, indian, indians, chinese, british end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	156 (57.35)	179 (55.76)	827 (71.23)	71.23
K-2 Words :	58 (21.32)	62 (19.31)	152 (13.09)	84.32
K-3 Words :	41 (15.07)	45 (14.02)	83 (7.15)	91.47
K-4 Words :	5 (1.84)	5 (1.56)	5 (0.43)	91.90
K-5 Words :	5 (1.84)	5 (1.56)	5 (0.43)	92.33
K-6 Words :	1 (0.37)	1 (0.31)	1 (0.09)	92.42
K-7 Words :	2 (0.74)	2 (0.62)	2 (0.17)	92.59
K-8 Words :				
K-9 Words :				
K-10 Words :				
K-11 Words :	1 (0.37)	1 (0.31)	1 (0.09)	92.68
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	2 (0.74)	2 (0.62)	6 (0.52)	93.20
K-16 Words :				
K-17 Words :	1 (0.37)	1 (0.31)	1 (0.09)	93.29
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	3 (0.93)	7 (0.60)	93.89
Total (unrounded)	272+?	321 (100)	1161 (100)	100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1161
Different words (types):	321
Type-token ratio:	0.28
Tokens per type:	3.62

Pertaining to onlist only

Tokens:	1154
Types:	318
Families:	272
Tokens per Family :	4.24
Types per Family :	1.17

Current profile (token %)	
K-1 (71.23)	71.23
K-2 (13.09)	84.32
K-3 (7.15)	91.47
K-4 (0.43)	91.90
K-5 (0.43)	92.33
K-6 (0.09)	92.42
K-7 (0.17)	92.59
K-11 (0.09)	92.68
K-15 (0.52)	93.20
K-17 (0.09)	93.29
OFF 0.60	100%

A. Types list

BNC-COCA-1,000 types: [fams 157 : types 180 : tokens 828]

a_[26] about_[2] across_[2] addition_[2] additional_[1] air_[1] all_[9] also_[7] among_[1] an_[7] and_[52] any_[2] are_[14] areas_[2] as_[15] at_[3] because_[4] become_[5] before_[1] besides_[1] better_[1] between_[2] billion_[1] both_[1] business_[2] but_[6] by_[5] bye_[2] can_[3] central_[2] computing_[1] continued_[1] control_[1] could_[1] count_[1] countries_[6]

definitely_1] difference_1] different_5] difficult_1] do_5] does_4]
education_3] either_2] especially_2] even_1] every_1] exactly_1]
fields_2] film_2] first_1] fit_1] five_1] for_10] from_2] further_1]
gives_1] giving_2] good_1] government_1] governmental_1] group_2]
growth_1] has_7] have_3] here_1] higher_1] historical_2] how_4]
hundreds_1] if_2] important_3] in_23] internet_1] into_2] is_39] it_20]
its_1] itself_1] job_1] know_3] knows_1] large_1] law_1] least_1]
less_1] level_1] little_1] living_1] long_1] look_1] major_2] making_1]
many_6] master_2] mastering_1] matter_1] mean_2] means_1]
million_3] more_6] most_8] much_1] music_2] nation_1] national_1]
nations_1] natural_1] new_2] nine_1] no_3] not_6] now_2] number_17]
numbers_3] of_57] on_1] one_4] only_5] open_1] or_12] other_3]
out_1] over_1] part_2] people_9] plays_2] position_1] power_6]
powerful_1] press_1] question_1] radio_1] read_1] really_3] reasons_1]
schools_1] science_1] sea_1] second_3] since_1] six_1] smaller_1]
so_4] some_3] south_1] speak_6] speakers_9] speaks_1] spoken_6]
sports_1] such_1] take_2] taught_1] television_1] than_3] thank_1]
that_10] the_86] their_1] them_1] there_8] they_3] third_1] this_4]
those_1] three_4] to_10] today_1] together_1] travel_1] true_1] try_1]
tv_1] two_1] types_1] uncertainty_1] up_1] use_7] used_7] user_1]
was_1] we_2] what_10] when_2] where_4] which_1] who_9] wide_2]
widely_5] with_4] words_1] world_8] written_2] yes_1] yet_2] you_7]

BNC-COCA-2,000 types: [fams 58 : types 62 : tokens 152]

account_1] backgrounds_1] commerce_1] common_2] commonly_1]
compare_1] competitions_1] contributes_1] development_1]
entertainment_1] events_1] example_1] finance_1] foreign_2] future_1]
include_1] increase_1] industries_1] industry_2] influence_2]
instance_1] introduction_1] language_56] languages_7] native_10]
non_2] official_6] opportunity_1] organizations_2] otherwise_1]
percent_2] plain_1] popular_2] population_5] practice_1] preferred_1]
presumably_1] previously_1] provider_1] qualify_1] range_1] regions_1]
represented_1] result_1] role_1] skills_1] spread_2] standard_1]
states_1] strength_1] strengthening_1] subcommittee_1] technology_1]
theatre_1] ticket_1] tongue_1] tongues_2] tourism_1] trade_1] traffic_1]
tricky_1] united_1] variants_1]

BNC-COCA-3,000 types: [fams 41 : types 45 : tokens 83]

academics_1] borders_1] categories_1] clarity_1] communicate_1]
communication_7] communications_1] continent_1] criteria_1]
defined_1] definition_1] distribution_2] diverse_1] dominant_4]
enables_1] estimated_2] estimates_2] extensively_1] extent_1]
facilitation_1] formerly_1] fulfils_1] furthermore_1] geographical_3]

global_8] institutions_1] international_12] justifies_1] label_1]
literature_1] majority_2] media_2] mutually_1] nevertheless_1]
poverty_1] precisely_1] primary_2] promotes_1] prospects_1] ratio_1]
sovereign_2] statistics_1] territories_1] universal_3] variety_1]

BNC-COCA-4,000 types: [fams 5 : types 5 : tokens 5]

competence_1] entities_1] faring_1] interface_1] steering_1]

BNC-COCA-5,000 types: [fams 5 : types 5 : tokens 5]

diplomacy_1] empowers_1] globe_1] intelligible_1] usage_1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

wofold_1]

BNC-COCA-7,000 types: [fams 2 : types 2 : tokens 2]

creole_1] butnumber_1]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

brevity_1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 2 : types 2 : tokens 6]

lingua_5] pidgin_1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens 1]

guesstimates_1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 7]

ethnicities_1] franca_5] natively_1]

B. Families list

BNC-COCA-1,000 Families: [fams 157 : types 180 : tokens 828]

a_33] about_2] across_2] add_3] air_1] all_9] also_7] among_1]
and_52] any_2] area_2] as_15] at_3] be_54] because_4] become_5]
before_1] beside_1] better_1] between_2] billion_1] both_1]
business_2] but_6] by_5] can_3] centre_2] certain_1] computer_1]
continue_1] control_1] could_1] count_1] country_6] definite_1]
difference_1] different_5] difficult_1] do_9] educate_3] either_2]
end_of_list_1] especially_2] even_1] every_1] exact_1] field_2] film_2]
first_1] fit_1] five_1] for_10] from_2] further_1] give_3] good_1]
goodbye_2] govern_2] group_2] grow_1] have_10] here_1] high_1]
history_2] how_4] hundred_1] if_2] important_3] in_23] internet_1]
into_2] it_22] job_1] know_4] large_1] law_1] least_1] less_1] level_1]
little_1] live_1] long_1] look_1] major_2] make_1] many_6] master_3]
matter_1] mean_3] million_3] more_6] most_8] much_1] music_2]
nation_3] nature_1] new_2] nine_1] no_3] not_6] now_2] number_20]
of_57] on_1] one_4] only_5] open_1] or_12] other_3] out_1] over_1]
part_2] people_9] play_2] position_1] power_7] press_1] question_1]

radio_1] read_1] really_3] reason_1] school_1] science_1] sea_1]
second_3] since_1] six_1] small_1] so_4] some_3] south_1] speak_22]
sport_1] such_1] take_2] teach_1] television_2] than_3] thank_1]
that_11] the_86] there_8] they_5] this_4] three_5] to_10] today_1]
together_1] travel_1] true_1] try_1] two_1] type_1] up_1] use_15]
we_2] what_10] when_2] where_4] which_1] who_9] wide_7] with_4]
word_1] world_8] write_2] yes_1] yet_2] you_7]

BNC-COCA-2,000 Families: [fams 58 : types 62 : tokens 152]

account_1] background_1] commerce_1] committee_1] common_3]
compare_1] competition_1] contribute_1] develop_1] entertain_1]
event_1] example_1] finance_1] foreign_2] future_1] include_1]
increase_1] industry_3] influence_2] instance_1] introduce_1]
language_63] native_10] non_2] official_6] opportunity_1] organize_2]
otherwise_1] percent_2] plain_1] popular_2] population_5] practise_1]
prefer_1] presume_1] previous_1] provide_1] qualify_1] range_1]
region_1] represent_1] result_1] role_1] skill_1] spread_2] standard_1]
states_1] strength_2] technology_1] theatre_1] ticket_1] tongue_3]
tour_1] trade_1] traffic_1] trick_1] unite_1] vary_1]

BNC-COCA-3,000 Families: [fams 41 : types 45 : tokens 83]

academy_1] border_1] category_1] clarify_1] communicate_9]
continent_1] criteria_1] define_2] distribute_2] diverse_1] dominant_4]
enable_1] estimate_4] extensive_1] extent_1] facilitate_1] former_1]
fulfil_1] furthermore_1] geography_3] global_8] institution_1]
international_12] justify_1] label_1] literature_1] majority_2] media_2]
mutual_1] nevertheless_1] poverty_1] precise_1] primary_2] promote_1]
prospect_1] ratio_1] sovereign_2] statistic_1] territory_1] universe_3]
variety_1]

BNC-COCA-4,000 Families: [fams 5 : types 5 : tokens 5]

competence_1] entity_1] fare_1] interface_1] steer_1]

BNC-COCA-5,000 Families: [fams 5 : types 5 : tokens 5]

diplomacy_1] empower_1] globe_1] intelligible_1] usage_1]

BNC-COCA-6,000 Families: [fams 1 : types 1 : tokens 1]

twofold_[1]

BNC-COCA-7,000 Families: [fams 2 : types 2 : tokens 2

creole_[1] outnumber_[1]

BNC-COCA-8,000 Families: [fams : types : tokens]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 1]

brevity_[1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 2 : types 2 : tokens 6]

lingua_[5] pidgin_[1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

guesstimate_[1]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 7]

ethnicities_[1] franca_[5] natively_[1]

7.2.3.3 *The Power of English Part 2*

7.2.3.3.1 Text only file

BEFORE YOU READ

What is the difference between colonialism and imperialism ?

Are there different kinds of imperialism ?

After the First World War the USA started to dominate the world politically , economically and culturally . The 20 century is often called the American Century .

The Power of English , Part 2

The power of English today is a consequence of the power of English in the past . Therefore , we have to unravel how English ended up at the top of the podium . How did English become the winner , a language spoken by only some 6 million people in the British Isles around the year 1600 ? The short answer to this question is power , exerted through imperialism , industrialization and trade .

The long answer takes us on a trip around the world and tells us how the British Empire expanded through settlements and trade . Other important aspects of the long answer are the Industrial Revolution that started in Britain and the growth of the USA as a super power .

The long answer starts in Ireland , England's first colony . Colonization of Ireland started early and gained ground when English and Scottish Protestants , called planters , settled in Ireland and became the ruling class . Their attempts at forcing their language , culture and religion on the Gaelic speaking Irish had far reaching and often bloody consequences , reverberating to this day .

Colonization made headway in the early 1600 with the settlement of the first islands in the Caribbean and the establishment of Jamestown in 1607 , in what later became the Colony of Virginia . This was the start of the first British Empire 1583 - 1783 . Gradually , more colonies were settled and claimed , both in the Caribbean and along the coast of North

" At one point the East India Company accounted for half of the world 's trade . True or false ?

America . By 1775 there were 20 British colonies in North America . Thirteen of these colonies joined forces to declare independence in 1776 .

After the War of Independence (1775 - 1783) , the United States of America became an independent nation , and English was the language of the government and the vast majority of the population . English was also firmly established in the remaining colonies in the

Caribbean and Canada , alongside the languages of the other colonial powers : French, Spanish and Dutch .

In the early 17 century the English East India Company founded trading posts in the East Indies , in what today are Indonesia, India, Pakistan and Bangladesh . These trading stations grew into settlements , and the trading companies , especially the East India Company , had enormous power , primarily economic power , but also political , judicial and military power . Competition for the lucrative trade in spices and tea was fierce , especially with the Netherlands and France, but Portugal and Spain also played important roles .

After Britain had lost its most populous colonies in America , the settlements in Asia became more important , and this shift in attention is considered the onset of the second British Empire (1783 - 1815) . During this period the British East India Company informally ruled India until 1858 , when the British government took over . Queen Victoria took the title Empress of India in 1876 , and British rule was formalized . By then English had become the language of administration , government and education in British India , a colony known as " the Jewel in the Crown " .

Britain also established outposts in other areas in Asia : Burma , Ceylon (now Sri Lanka) , Malaya and British Borneo (both now part of Malaysia) , Brunei , Singapore , and Hong Kong , all of which came under British rule after 1815 . Trade was the corner stone of British imperialism , and the British merchant fleet and Navy were present in all parts of the world . Consequently , British power and the English language seeped into many different nooks and crannies .

By 1770 James Cook had claimed both Australia and New Zealand for the British crown . Some 160 000 British convicts were transported to Australia between 1788 and 1868 . They were joined by free immigrants , but the sea voyage was expensive and hazardous , so their numbers stayed relatively low . During the Gold Rush , starting in 1851 , immigration increased , and large numbers of British and Irish settlers tried their luck in Australia . There were of course also immigrants from other European and Asian countries , and by 1900 the total immigrant population was nearly 4 million . After the Second World War Australia launched a mass immigration program , and immigrants from the UK and Ireland far outnumbered immigrants from all other countries . New Zealand was settled around the same

time as Australia , but the official colony was not established until 1840 . From then on , immigration picked up , and by 1911 there were a million European immigrants , primarily British and Irish . Thus , English came to be the official language of this part of the world .

Queen Victoria (1819 - 1901) and her Indian secretary at Balmoral Castle in Royal Deeside in Aberdeenshire , Scotland . She was Queen of Great Britain and Ireland (1837 - 1901) and Empress of India (1876 - 1901) .

Adventurous English merchants started trading in West Africa as early as 1530 , but the first English fort was not set up until 1663 in Gambia .

England and other European colonial powers built coastal forts as a base for trade , which developed into areas of influence along the coast strips . The vast interior of Africa was not colonized and was little known to the Europeans until the late 18 century . Then the "

Scramble for Africa " began , and by 1913 , 90 percent of Africa was under European control . Yet there was never any mass emigration from Britain to Africa .

Until 1865 the slave trade was the most important business in West Africa , and millions of African slaves were transported to the Americas under the most barbaric conditions . Britain was also interested in Egypt and South Africa in order to maintain secure transportation routes to India . British colonialists wanted to build a railway from Cairo to the Cape of Good Hope . In this process they occupied and colonized the Cape Colony , which later became a part of South Africa . In Egypt the Suez Canal was the most important asset and therefore protected by British soldiers . Egypt was a British protectorate from 1914 until 1922 , but it was never a colony .

By the mid 1800s the Industrial Revolution had transformed England , and the booming industries needed raw materials , especially from Africa . The expanding global trade boosted the British economy and gave Britain even more economic and industrial power , in addition to technological and scientific power . The Empire had given Britain political and military power and , combined with economic and cultural power , this explains the background for the growth of English as a global language .

The year 1815 was a turning point in European history because Napoleon and France were finally defeated by a coalition of European armies . Britain benefited from the peace treaties , and this marked the start of the British Imperial Century (1815-1914). Britain was the leading global power all over the world, and English was the language of the people in power . To describe the expansion of the British Empire , the phrase " The Empire on which the sun never sets " was used , because the sun was always shining in some territory of the Empire .

Around 1922 the British Empire was at its peak . It covered almost a quarter of all land on Earth and ruled over 458 million people , more than

one fifth of the world's population at the time . It goes without saying that this resulted in an immense influence on the political and legal systems and cultures in all regions belonging to the Empire . It also resulted in the unrivalled power of the English language .

By the beginning of the 20 century the United States and Germany had begun to seriously challenge Britain's position as " Work shop of the

World " . After the First World War , the USA was superior in terms of military and industrial power . From then on , the USA started to dominate the world in many different ways , and the 20 century is often called the American Century . As the British Empire was gradually dismantled and colonies

gained independence , its political and military power was reduced . Its position was taken over by the USA , but they ruled the world with a different kind of imperialism .

In the course of the 20 century the USA became an economic super power . As the world's largest economy in terms of GDP , the USA has both industrial and financial power . There is also technological superiority ; the USA being the world leader in developing new technologies , even though other countries are catching up in this field . In addition , there is hard power : military and political power , and soft power : ideological and cultural power . American soldiers and weapons have played important roles in many international conflicts during the 20 century . How this role will develop in What kind of power did the USA have in

the 20 century ? How did the power of the United States contribute to the consolidation of the position of English as

a global language ?

Translation is a tricky business . What do you think the correct English phrase is ?

the 21 century , after the Iraq and Afghan Wars , remains to be seen . American ideas and ideals and American popular culture and life style remain highly influential . They are visible around the world through literature, music, films , television series , social media , and other means .

The English of the 20 century was most definitely American English , and its voice was heard all around the world . Stop and think for a moment : are you not exposed to American English in some form or other every single day ?

So far this text has explained how the English language has spread around the world as a result of two main factors : the UK and the USA . First , there was the colonialism and imperialism of the British Empire in the 17 , 18 and 19 centuries . In the 20 century there was the supremacy of the USA . This also explains to some extent why English ended up as the winner .

Yet it does not fully explain why English won and not , for example , Spanish or Chinese . That is a complex issue which we cannot discuss here . Suffice it is to say , in the words of the linguist David Crystal , that English " is a language which has repeatedly found itself in the right place at the right time " . In the 20 century , as international developments brought about the need for a lingua franca , English was a clear first choice . In addition , English was also the first choice of many nations that became independent in the 20 century and needed a neutral administrative language . Then , finally ,there was the digital revolution , where English too was in the right place at the right time . It has been said , with more than a little irony, that English as a global language might have suffered a set back if Bill Gates had grown up speaking Chinese . Triumph of Steam and Electricity . An English lithograph of 1897 which illustrates scenes of scientific progress during Queen Victoria's reign . Scenes from 1897 are on the left and scenes from 1937 are on the right . Describe each pair of scenes . Can you identify the scientists around the border ?

7.2.3.3.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

one-fifth, Gaelic-speaking, far-reaching,

4. Compound words separated:

superpower, workshop, lifestyle, cornerstone, setback, railway

5. Words (groups of letters) removed from the text analysis:

“th”, (17, 19, 20th), “st” (21st),

6. Proper nouns:

USA, American, English, British, Britain, USA, Ireland, England, Scottish, Gaelic, Caribbean, Jamestown, Virginia, America, India, Indies, Indonesia, Pakistan, Bangladesh, Netherlands, France, Portugal, Spain, Asia, Victoria, Burma, Ceylon, Sri Lanka, Malaya, Malaysia, Borneo, Brunei, Singapore, Hong Kong, James Cook, Australia, Zealand, British, Irish, UK, European, Indian, Balmoral, Deeside, Aberdeenshire, Scotland, Africa, European, Europeans, Americas, Egypt, Cairo, Suez, Napoleon, France, Germany, GDP, Iraq, Afghan, David, Crystal, Bill, Gates, Chinese, Asian, French, Spanish, Dutch, Canada, Gambia, African, Protestants,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

First World War, (American) Century, (British) Isles, (British) Empire, (Industrial Revolution), (Scottish Protestants), (Gaelic)-speaking (Irish), Colony of (Virginia), North (America), East (India) Company, (War of Independence), United States of (America), Canada, French, Spanish, Dutch, East (Indies), Queen (Victoria), Empress of (India), (British India), Jewel in the Crown, British (Borneo), (Hong Kong), (British) Navy, New (Zealand), Gold Rush, Second World War, (Balmoral) Castle, Royal (Deeside), Queen of Great (Britain), West (Africa), Gambia, Scramble for (Africa), South (Africa), Cape of Good Hope, Cape Colony, Industrial Revolution, Earth, United States, Workshop of the World, First World War, (American) Century, (Iraq) War, (Afghan) Wars, (Suez) Canal, The Triumph of Steam and Electricity

Note: Text related to illustrations have been included in the text analysis.

7.2.3.3.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: The Power of English Part 2 (11.37 kb)

Words recategorized by user as 1k items (proper nouns etc): USA, AMERICAN, ENGLISH, BRITISH, BRITAIN, USA, IRELAND, ENGLAND, SCOTTISH, GAELIC, CARIBBEAN, JAMESTOWN, VIRGINIA, AMERICA, INDIA, INDIES, INDONESIA, PAKISTAN, BANGLADESH, NETHERLANDS, FRANCE, PORTUGAL, SPAIN, ASIA, VICTORIA, BURMA, CEYLON, SRI LANKA, MALAYA, MALAYSIA, BORNEO, BRUNEI, SINGAPORE, HONG KONG, JAMES COOK, AUSTRALIA, ZEALAND, BRITISH, IRISH, UK, EUROPEAN, INDIAN, BALMORAL, DEESIDE, ABERDEENSHIRE, SCOTLAND, AFRICA, EUROPEAN, EUROPEANS, AMERICAS, EGYPT, CAIRO, SUEZ, NAPOLEON, FRANCE, GERMANY, GDP, IRAQ, AFGHAN, DAVID, CRYSTAL, BILL, GATES, CHINESE, ASIAN, FRENCH, SPANISH, DUTCH, CANADA, GAMBIA, AFRICAN, PROTESTANTS, (total 202 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	277	352	1452	77.81%
Function:	(799)	(42.82%)
Content:	(653)	(34.99%)

	> Anglo-Sax =Not Greco-Lat/Fr Cog:	(299)	(16.02%)
K2 Words (1001-2000):		46	49	74	3.97%
	> Anglo-Sax:	(15)	(0.80%)
1k+2k		(81.78%)
AWL Words (academic):		51	66	97	5.20%
	> Anglo-Sax:	(1)	(0.05%)
Off-List Words:		?	132	243	13.02%
		374+?	599	1866	100%

Words in text (tokens):	1866
Different words (types):	599
Type-token ratio:	0.32
Tokens per type:	3.12
Lex density (content words/total)	0.57

Pertaining to onlist only

Tokens:	1623
Types:	467
Families:	374
Tokens per family:	4.34
Types per family:	1.25
Anglo-Sax Index: (A-Sax tokens + functors / onlist tokens)	%
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
77.81	77.81
3.97	81.78
5.20	86.98
13.02	100.00

A. AWL Tokens lists

AWL [51:66:97] administration administrative areas areas aspects benefited challenge complex conflicts consequence consequences consequently contribute cultural cultural culturally culture culture cultures definitely dominate dominate economic economic economic economic economically economy economy enormous established established established

establishment expanded expanding expansion exposed factors finally finally financial founded global global global global global identify ideological illustrates immigrant immigrants immigrants immigrants immigrants immigrants immigration immigration immigration issue legal maintain majority media military military military military military neutral occupied percent period primarily primarily process regions revolution revolution revolution role roles roles routes secure series shift style technological technological text transformed transportation transported transported visible

Sublist 1

areas areas benefited economic economic economic economic economically economy economy established established established establishment factors financial identify issue legal majority percent period process role roles roles

Sublist 2

administration administrative aspects complex consequence consequences consequently cultural cultural culturally culture culture cultures finally finally maintain primarily primarily regions secure text

Sublist 3

contribute dominate dominate illustrates immigrant immigrants immigrants immigrants immigrants immigrants immigration immigration immigration shift technological technological

Sublist 4

occupied series

Sublist 5

challenge conflicts expanded expanding expansion exposed style

Sublist 6

neutral transformed transportation transported transported

Sublist 7

definitely global global global global global ideological media visible

Sublist 9

founded military military military military military revolution revolution revolution routes

Sublist 10

enormous

B. AWL Types list

AWL types: [51:66:97] administration_[1] administrative_[1] areas_[2] aspects_[1] benefited_[1] challenge_[1] complex_[1] conflicts_[1]

consequence [1] consequences [1] consequently [1] contribute [1]
 cultural [2] culturally [1] culture [2] cultures [1] definitely [1] dominate [2]
 economic [4] economically [1] economy [2] enormous [1] established [3]
 establishment [1] expanded [1] expanding [1] expansion [1] exposed [1]
 factors [1] finally [2] financial [1] founded [1] global [5] identify [1]
 ideological [1] illustrates [1] immigrant [1] immigrants [5] immigration [3]
 issue [1] legal [1] maintain [1] majority [1] media [1] military [5]
 neutral [1] occupied [1] percent [1] period [1] primarily [2] process [1]
 regions [1] revolution [3] role [1] roles [2] routes [1] secure [1] series [1]
 shift [1] style [1] technological [2] text [1] transformed [1]
 transportation [1] transported [2] visible [1]

C. AWL Families list

AWL families: [51:66:97]

administer [2] area [2] aspect [1] benefit [1] challenge [1] complex [1]
 conflict [1] consequent [3] contribute [1] **culture** [6] definite [1]
 dominate [2] **economy** [7] enormous [1] establish [4] expand [3]
 expose [1] factor [1] final [2] finance [1] founded [1] globe [5] identify [1]
 ideology [1] illustrate [1] **immigrate** [9] issue [1] legal [1] maintain [1]
 major [1] media [1] military [5] neutral [1] occupy [1] percent [1]
 period [1] primary [2] process [1] region [1] revolution [3] role [3]
 route [1] secure [1] series [1] shift [1] style [1] technology [2] text [1]
 transform [1] transport [3] visible [1]

AWL Fr non-cognate families: [families 1 : tokens 1] shift [1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, probers, cognates, extraction, barchart)

WEB VP OUTPUT FOR FILE: The Power of English Part 2 (11,690 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): usa, american, english, british, britain, usa, ireland, england, scottish, gaelic, caribbean, jamestown, virginia, america, india, indies, indonesia, pakistan, bangladesh, netherlands, france, portugal, spain, asia, victoria, burma, ceylon, sri , lanka, malaya, malaysia, borneo, brunei, singapore, hong, kong, james, cook, australia, zealand, british, irish, uk, european, indian, balmoral, deeside, aberdeenshire, scotland, africa, european, europeans, americas, egypt, cairo, suez, napoleon, france, germany, gdp, iraq, afghan, david, crystal, bill, gates, chinese, asian, french, spanish, dutch, canada, gambia, african, protestants, end_of_list

Cognates => 1k: None

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as

program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	312 (62.28)	368 (61.33)	1488 (79.74)	79.74
K-2 Words :	86 (17.17)	109 (18.17)	201 (10.77)	90.51
K-3 Words :	63 (12.57)	75 (12.50)	114 (6.11)	96.62
K-4 Words :	15 (2.99)	17 (2.83)	19 (1.02)	97.64
K-5 Words :	5 (1.00)	5 (0.83)	5 (0.27)	97.91
K-6 Words :	5 (1.00)	5 (0.83)	5 (0.27)	98.18
K-7 Words :	3 (0.60)	3 (0.50)	3 (0.16)	98.34
K-8 Words :	4 (0.80)	4 (0.67)	5 (0.27)	98.61
K-9 Words :	4 (0.80)	4 (0.67)	4 (0.21)	98.82
K-10 Words :	1 (0.20)	1 (0.17)	1 (0.05)	98.87
K-11 Words :	2 (0.40)	2 (0.33)	2 (0.11)	98.98
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	1 (0.20)	1 (0.17)	1 (0.05)	99.03
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	2 (0.33)	2 (0.11)	99.14
Total (unrounded)	501+?	600 (100)	1866 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens): 1866
 Different words (types): 600
 Type-token ratio: 0.32
 Tokens per type: 3.11

Pertaining to onlist only

Tokens: 1864
 Types: 598
 Families: 501
 Tokens per Family : 3.72
 Types per Family : 1.19

Current profile (token %)	
K-1 (79.74)	79.74
K-2 (10.77)	90.51
K-3 (6.11)	96.62
K-4 (1.02)	97.64
K-5 (0.27)	97.91
K-6 (0.27)	98.18
K-7 (0.16)	98.34
K-8 (0.27)	98.61
K-9 (0.21)	98.82
K-10 (0.05)	98.87
K-11 (0.11)	98.98
K-15 (0.05)	99.03
OFF (0.11)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 222 : types 268 : tokens 1305]

a [29] about [1] addition [3] after [7] all [7] almost [1] along [2] also [10]
 always [1] an [4] and [91] answer [4] any [1] are [8] areas [2] around [8]
 as [15] at [8] **back** [1] base [1] be [2] became [7] because [2]
 become [2] been [1] before [1] began [1] beginning [1] begun [1]
 being [1] between [2] bill [1] bloody [1] both [4] brought [1] build [1]
 built [1] business [2] but [7] by [13] called [3] came [2] can [1]
 cannot [1] catching [1] choice [2] class [1] clear [1] companies [1]
 company [4] considered [1] control [1] cook [1] **corner** [1] countries [3]
 course [2] covered [1] day [2] definitely [1] did [3] difference [1]
 different [4] do [1] does [1] during [4] each [1] early [4] earth [1] east [5]

education [1] ended [2] especially [3] even [2] every [1] expensive [1]
 explain [1] explained [1] explains [2] far [3] field [1] fifth [1] films [1]
 finally [2] first [9] for [9] forces [1] forcing [1] form [1] found [1] free [1]
 from [12] fully [1] gave [1] given [1] goes [1] gold [1] good [1]
 government [3] great [1] grew [1] ground [1] grown [1] growth [2] had [9]
 half [1] hard [1] has [5] have [4] heard [1] her [1] here [1] highly [1]
 history [1] hope [1] how [6] ideas [1] if [1] important [6] in [64]
 interested [1] into [3] is [13] islands [1] issue [1] it [7] its [5] itself [1]
 joined [2] kind [2] kinds [1] known [2] land [1] large [1] largest [1]
 late [1] later [2] leader [1] leading [1] left [1] life [1] little [2] long [3]
 lost [1] low [1] luck [1] made [1] main [1] many [4] marked [1] means [1]
 might [1] million [4] millions [1] moment [1] more [5] most [5] music [1]
 nation [1] nations [1] nearly [1] need [1] needed [2] never [3] new [3]
 north [2] not [6] now [2] number [62] numbers [3] of [69] often [3] on [9]
 one [2] only [1] or [3] order [1] other [9] over [4] pair [1] part [4]
 parts [1] past [1] people [3] picked [1] place [2] planters [1] played [2]
 point [2] position [3] **posts** [1] power [26] powers [2] present [1]
 program [1] protected [1] quarter [1] queen [4] question [1] **teaching** [1]
 read [1] right [5] rule [2] ruled [3] ruling [1] said [1] same [1] say [1]
 saying [1] scientific [2] scientists [1] sea [1] second [2] secure [1]
 seen [1] seriously [1] set [2] sets [1] **settled** [3] **settlement** [1]
settlements [3] **settlers** [1] she [1] shop [1] short [1] single [1] so [2]
 soft [1] some [5] south [2] speaking [2] spoken [1] sri [1] start [2]
 started [5] starting [1] starts [1] stations [1] stayed [1] stone [1] stop [1]
 sun [2] systems [1] taken [1] takes [1] tea [1] television [1] tells [1]
 terms [2] than [2] that [6] the [183] their [4] then [5] there [10] these [2]
 they [4] think [2] thirteen [1] this [14] though [1] through [3] time [4]
 to [24] today [2] too [1] took [2] top [1] total [1] tried [1] trip [1] true [1]
 turning [1] two [1] under [3] until [6] up [6] us [2] used [1] voice [1]
 wanted [1] war [4] wars [1] was [39] ways [1] we [2] were [9] west [2]
 what [5] when [2] where [1] which [7] why [2] will [1] winner [2] with [5]
 without [1] won [1] words [1] **work** [1] world [18] year [2] yet [2] you [4]
 *part of compound words in the glossary: workshop, cornerstone, setback, far-reaching

BNC-COCA-2,000 types: [fams 87 : types 102 : tokens 202]

accounted [1] **adventurous** [1] armies [1] **attempts** [1] attention [1]
 background [1] belonging [1] benefited [1] **booming** [1] cape [2]
 castle [1] centuries [1] century [16] challenge [1] **claimed** [2] coast [2]
 coastal [1] combined [1] competition [1] conditions [1] contribute [1]
 correct [1] crown [2] cultural [2] culturally [1] culture [2] cultures [1]
 describe [2] develop [1] developed [1] developing [1] developments [1]
 discuss [1] economic [4] economically [1] economy [2] electricity [1]
 empire [11] enormous [1] established [3] establishment [1] example [1]

exposed [1] financial [1] firmly [1] gained [2] gates [1] identify [1]
illustrates [1] increased [1] industrial [5] industrialization [1] industries [1]
influence [2] influential [1] language [14] languages [1] legal [1]
maintain [1] mass [2] materials [1] military [5] navy [1] official [2]
peace [1] percent [1] period [1] political [5] politically [1] popular [1]
population [3] process [1] progress [1] reduced [1] regions [1] remain [1]
remaining [1] remains [1] repeatedly [1] result [1] resulted [2] role [1]
roles [2] royal [1] rush [1] scenes [4] series [1] shift [1] shining [1]
slave [1] slaves [1] social [1] soldiers [2] spread [1] states [3] steam [1]
strips [1] style [1] suffered [1] super [2] technological [2] technologies [1]
therefore [2] thus [1] title [1] trade [8] trading [4] tricky [1] united [3]
weapons [1]

BNC-COCA-3,000 types: [fams 64 : types 73 : tokens 115]

administration [1] administrative [1] alongside [1] aspects [1] asset [1]
boosted [1] border [1] coalition [1] colonial [2] colonialism [2]
colonialists [1] colonies [6] colonization [2] colonized [2] colony [6]
complex [1] conflicts [1] consequence [1] consequences [1]
consequently [1] convicts [1] crystal [1] declare [1] defeated [1] digital [1]
dominate [2] expanded [1] expanding [1] expansion [1] extent [1]
factors [1] false [1] fierce [1] formalized [1] founded [1] global [5]
gradually [2] hazardous [1] ideals [1] immigrant [1] immigrants [5]
imperial [1] imperialism [6] independence [3] independent [2] interior [1]
international [2] irony [1] launched [1] literature [1] majority [1] media [1]
neutral [1] occupied [1] peak [1] phrase [2] primarily [2] raw [1]
relatively [1] religion [1] revolution [3] routes [1] secretary [1] superior [1]
superiority [1] territory [1] text [1] transformed [1] translation [1]
transportation [1] transported [2] treaties [1] triumph [1] unrivalled [1]
vast [2] visible [1]

BNC-COCA-4,000 types: [fams 15 : types 17 : tokens 19]

canal [1] consolidation [1] exerted [1] fleet [1] fort [1] forts [1]
ideological [1] immense [1] immigration [3] informally [1] judicial [1]
merchant [1] merchants [1] mid [1] reign [1] scramble [1] spices [1]

BNC-COCA-5,000 types: [fams 5 : types 5 : tokens 5]

dismantled [1] emigration [1] jewel [1] onset [1] voyage [1]

BNC-COCA-6,000 types: [fams 5 : types 5 : tokens 5]

isles [1] lucrative [1] scepers [1] suffice [1] unravel [1]

BNC-COCA-7,000 types: [fams 3 : types 3 : tokens 3]

outnumbered [1] reverberating [1] supremacy [1]

BNC-COCA-8,000 types: [fams 4 : types 4 : tokens 5]

empress [2] linguist [1] outposts [1] podium [1]

BNC-COCA-9,000 types: [fams 4 : types 4 : tokens 4]

barbaric [1] headway [1] lithograph [1] nooks [1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

populous [1]

BNC-COCA-11,000 types: [fams 2 : types 2 : tokens 2]

crannies [1] protectorate [1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams 1 : types 1 : tokens 1]

bangladesh [1]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 1]

lingua [1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

franca [1] railway [1]

B. Families list

BNC-COCA-1,000 Families: [fams 222 : types 268 : tokens 1305]

a [33] about [1] add [3] after [7] all [7] almost [1] along [2] also [10] always [1] and [91] answer [4] any [1] area [2] around [8] as [15] at [8] back [1] base [1] be [73] because [2] become [9] before [1] begin [3] between [2] bill [1] blood [1] both [4] bring [1] build [2] business [2] but [7] by [13] call [3] can [2] catch [1] choice [2] class [1] clear [1] come [2] company [5] consider [1] control [1] cook [1] corner [1] country [3] course [2] cover [1] day [2] definite [1] difference [1] different [4] do [5] during [4] each [1] early [4] earth [1] east [5] educate [1] end [2] especially [3] even [2] every [1] expensive [1] explain [4] far [3] field [1] film [1] final [2] find [1] first [9] five [1] for [9] force [2] form [1] free [1] from [12] full [1] give [2] go [1] gold [1] good [1] govern [3] great [1] ground [1] grow [4] half [1] hard [1] have [18] hear [1] here [1] high [1] history [1] hope [1] how [6] idea [1] if [1] important [6] in [64] interest [1] into [3] island [1] issue [1] it [13] join [2] kind [3] know [2] land [1] large [2] late [3] lead [2] left [1] life [1] little [2] long [3] lose [1] low [1] luck [1] main [1] make [1] many [4] mark [1] mean [1] might [1] million [5] moment [1] more [5] most [5] music [1] nation [2] near [1] need [3] never [3] new [3] north [2] not [6] now [2] number [65] of [69] often [3] on [9] one [2] only [1] or [3] order [1] other [9] over [4] pair [1] part [5] past [1] people [3] pick [1] place [2] plant [1] play [2] point [2] position [3] post [1] power [28] present [1] programme [1] protect [1] quarter [1] queen [4] question [1] reach [1] read [1] right [5] rule [6] same [1] say [3] science [3] sea [1] second [2] secure [1] see [1] serious [1] set [3] settle [8] she [2] shop [1] short [1] single [1] so [2] soft [1] some [5] south [2] speak [3] start [9] station [1] stay [1] stone [1] stop [1] sun [2] system [1] take [4] tea [1] television [1] tell [1] term [2] than [2] that [6] the [183] then [5] there [10] they [8] think [2] thirteen [1] this [16] though [1] through [3] time [4] to [24] today [2] too [1] top [1] total [1] trip [1] true [1] try [1] turn [1] two [1] under [3] until [6] up [6] use [1] voice [1] want [1] war [5] way [1] we [4] west [2] what [5] when [2] where [1] which [7]

why [2] will [1] win [3] with [5] without [1] word [1] work [1] world [18] year [2] yet [2] you [4]

BNC-COCA-2,000 Families: [fams 87 : types 102 : tokens 202]

account [1] adventure [1] army [1] attempt [1] attention [1] background [1] belong [1] benefit [1] boom [1] cape [2] castle [1] century [17] challenge [1] claim [2] coast [3] combine [1] competition [1] condition [1] contribute [1] correct [1] crown [2] culture [6] describe [2] develop [4] discuss [1] economy [7] electric [1] empire [11] enormous [1] establish [4] example [1] expose [1] finance [1] firm [1] gain [2] gate [1] identify [1] illustrate [1] increase [1] industry [7] influence [3] language [15] legal [1] maintain [1] mass [2] material [1] military [5] navy [1] official [2] peace [1] percent [1] period [1] politics [6] popular [1] population [3] process [1] progress [1] reduce [1] region [1] remain [3] repeat [1] result [3] role [3] royal [1] rush [1] scene [4] series [1] shift [1] shine [1] slave [2] social [1] soldier [2] spread [1] states [3] steam [1] strip [1] style [1] suffer [1] super [2] technology [3] therefore [2] thus [1] title [1] trade [12] trick [1] unite [3] weapon [1]

BNC-COCA-3,000 Families: [fams 64 : types 73 : tokens 115]

administration [1] administrative [1] alongside [1] aspect [1] asset [1] boost [1] border [1] coalition [1] colony [21] complex [1] conflict [1] consequence [2] consequent [1] convict [1] crystal [1] declare [1] defeat [1] digital [1] dominate [2] expand [2] expansion [1] extent [1] factor [1] false [1] fierce [1] formal [1] founded [1] global [5] gradual [2] hazard [1] ideal [1] immigrant [6] imperial [7] independence [3] independent [2] interior [1] international [2] irony [1] launch [1] literature [1] majority [1] media [1] neutral [1] occupy [1] peak [1] phrase [2] primary [2] raw [1] relative [1] religion [1] revolution [3] rival [1] route [1] secretary [1] superior [2] territory [1] text [1] transform [1] translate [1] transport [3] treaty [1] triumph [1] vast [2] visible [1]

BNC-COCA-4,000 Families: [fams 15 : types 17 : tokens 19]

canal [1] consolidate [1] exert [1] fleet [1] fort [2] ideological [1] immense [1] immigrate [3] informal [1] judicial [1] merchant [2] mid [1] reign [1] scramble [1] spice [1]

BNC-COCA-5,000 Families: [fams 5 : types 5 : tokens 5]

dismantle [1] emigrate [1] jewel [1] onset [1] voyage [1]

BNC-COCA-6,000 Families: [fams 5 : types 5 : tokens 5]

isle_[1] lucrative_[1] seep_[1] suffice_[1] unravel_[1]

BNC-COCA-7,000 Families: [fams 3 : types 3 : tokens 3]

outnumber_[1] reverberate_[1] supremacy_[1]

BNC-COCA-8,000 Families: [fams 4 : types 4 : tokens 5]

empress_[2] linguist_[1] outpost_[1] podium_[1]

BNC-COCA-9,000 Families: [fams 4 : types 4 : tokens 4]

barbaric_[1] headway_[1] lithograph_[1] nook_[1]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

populous_[1]

BNC-COCA-11,000 Families: [fams 2 : types 2 :

cranny_[1] protectorate_[1]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams 1 : types 1 : tokens

bangladesh_[1]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens

lingua_[1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

franca_[1] railway_[1]

7.2.3.4 *English and the Future*

7.2.3.4.1 Text only file

ENGLISH AND THE FUTURE

What is the future of English ? How will language learning change ? Will students learn in the same way in the future ? What will be the biggest change ?

To find out some answers to these questions , we asked David Graddol , author of ' The Future of English ? ' , to give us his views . Read his article below and then join in the discussion at the bottom of this page .

Introducing David Graddol

David Graddol

David Graddol is a British applied linguist , writer , broadcaster , researcher and consultant on global English . David wrote a follow up analysis of global trends in English language learning - ' English Next ' - which was published by the British Council in February 2006 .

The Article - Learners of the future

A fast train in Tokyo - picture by Carlos

The world is changing so fast that English , perhaps the most worldly of languages , is struggling to keep up . One thing is for sure : the English learner of the future will be different from those of the past , will be looking for a different kind of English and will expect to learn it in ways which reflect the technology and life styles of the 21 century .

Learners in the future are likely to be much younger . Across the world , English is being made a central component of more general educational reform . English is losing its position in the foreign languages curriculum , where it was taught mainly to teenagers and has been reinvented as one of the basic skills which you need to learn when you first go to school . Text books and audio visual materials , methods of teaching and expected outcomes are already being transformed .

Children in Egypt - picture by Shaden

Young Learners

Young children are often said to be better at language learning than older learners but they also have special challenges . Young children do not usually have the kind of instrumental

motivation and determination for learning English that older learners often have (though their parents and relations may) . English lessons must therefore be fun and rewarding . Young learners also have less experience at learning and so fewer cognitive strategies for remembering things , or coping with the discouraging set backs that are typical of any learning curve . Highly visual web sites with interactive games which rely less on written text will provide accessible support for such learners .

As General English becomes something done when you are young , teenagers and young adults will be seeking more specific needs and knowledge areas . In fact , one of the consequences of the universalisation of English is the convergence between knowledge , skills and English . So learning about anything in future - whether computers or football - may come with an element of specialised English learning .

The countries where English is most sought after are also changing . As developing economies and growing populations create more demand for English , the global class room is getting ever fuller . Learners from Brazil , Poland and China are joining class mates from Japan and Korea . But the internet is also supporting many minority learners .

People in Brazil - picture by Denis
Why Learn English ?

The reasons why people learn English are also changing . Globalisation is bringing together more people than ever who speak different languages and who are turning to English as the means of communication . The English learner of the future may be less worried about sounding exactly like a native speaker and more concerned about how to use English effectively in cross cultural communication . We may be hearing more non native speakers in dialogues and a wider range of the ' New Englishes ' now used around the world .

Technology will allow English to come to you , rather than you having to go to a special place to learn English . Podcasts and downloadable computer programs hint at the range of things to come as the distinction between televisions , computers , mobile phones and mp3 players gets more blurred . And it is not just the technologies which are converging ? it also increasingly difficult to tell the difference between providers of educational content , service providers and hardware manufacturers . That may be one reason why support for learning English is coming from an increasing number of sources .

Learning English has always involved both pain and pleasure , private slog and social activity . Traditional learning provided take it or leave it mixes of these as well as of content but in future learners will expect be able to choose a formula which suits their cultural and psychological dispositions , or their particular needs at that moment . They , rather than their teachers , will decide how , what and when they will learn . Web sites will provide the kind of support needed by learners to chart a pathway through the material and monitor progress .

People in India - picture by Sammay
Communication

Above all , learning English is about communication and an important part of learning English is being able to exchange views and make friends with people all over the world . As learners become younger , this has a dark side as well . Issues of security and transparency of

identity will become greater . Despite the growing independence of learners , trusted institutions and brand names will remain important .

Lastly , in envisioning the learners of the future as younger and more demanding , it is also worth considering the teachers of the future . The paradox is , as English becomes spoken by more and more people in the world , the number of English teachers will fall .

7.2.3.4.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

don't, you're

3. Hyphenated words with hyphen removed:

follow-up, cross-cultural, non-native, take-it, leave-it

4. Compound words separated:

Lifestyles, classroom, classmates, textbook, setback, website,

5. Words (groups of letters) removed from the text analysis:

21 (st)

6. Proper nouns:

English, David, Graddol, British, February, Tokyo, Carlos, Egypt, Shaden, Brazil , Poland, China, Japan, Korea, Denis, Englishes, India, Sammay,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

(British) Council, General (English), New (Englishes),

Note: Text related to illustrations have been included in the text analysis.

7.2.3.4.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Targets - English and the Future (5.55 kb)

Words recategorized by user as 1k items (proper nouns etc): ENGLISH, DAVID, GRADDOL, BRITISH, FEBRUARY, TOKYO, CARLOS, EGYPT, SHADEN, BRAZIL , POLAND, CHINA, JAPAN, KOREA, DENIS, ENGLISHES, INDIA, SAMMAY (total 58 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	217	264	749	83.59%

Function:	(394)	(43.97%)
Content:	(355)	(39.62%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	(219)	(24.44%)
K2 Words (1001-2000):	21	21	23	2.57%
> Anglo-Sax:	(9)	(1.00%)
1k+2k	(86.16%)
AWL Words (academic):	48	51	63	7.03%
> Anglo-Sax:	(7)	(0.78%)
Off-List Words:	?	51	61	6.81%
	286+?	387	896	100%

Words in text (tokens):	896
Different words (types):	387
Type-token ratio:	0.43
Tokens per type:	2.32
Lex density (content words/total)	0.56

Pertaining to onlist only

Tokens:	835
Types:	336
Families:	286
Tokens per family:	2.92
Types per family:	1.17
Anglo-Sax Index: <small>(A-Sax tokens + functors / onlist tokens)</small>	%
Greco-Lat/Fr-Cognate Index: <small>(Inverse of above)</small>	%

Current profile	
%	Cumul.
83.59	83.59
2.57	86.16
7.03	93.19
6.81	100.00

A. AWL Tokens lists

AWL [48:51:63] accessible adults analysis areas author challenges chart communication communication communication communication component

computer computers computers consequences consultant create cultural cultural despite distinction economies element formula global global global globalisation identity institutions interactive involved issues methods minority monitor motivation outcomes psychological published range range rely researcher security seeking sites sites sought sources specific strategies styles technology technology text text traditional transformed trends visual visual

Sublist 1

analysis areas create economies formula identity involved issues methods researcher sources specific

Sublist 2

computer computers computers consequences cultural cultural distinction element institutions range range range security seeking sites sites sought strategies text text traditional

Sublist 3

component interactive minority outcomes published rely technology technology

Sublist 4

accessible communication communication communication communication despite

Sublist 5

challenges consultant monitor psychological styles trends

Sublist 6

author motivation transformed

Sublist 7

adults global global global globalisation

Sublist 8

chart visual visual

B. AWL Types list

AWL types: [48:51:63] accessible_[1] adults_[1] analysis_[1] areas_[1] author_[1] challenges_[1] chart_[1] communication_[4] component_[1] computer_[1] computers_[2] consequences_[1] consultant_[1] create_[1] cultural_[2] despite_[1] distinction_[1] economies_[1] element_[1] formula_[1] global_[3] globalisation_[1] identity_[1] institutions_[1] interactive_[1] involved_[1] issues_[1] methods_[1] minority_[1] monitor_[1]

motivation_1] outcomes_1] psychological_1] published_1] range_2]
 rely_1] researcher_1] security_1] seeking_1] sites_2] sought_1]
 sources_1] specific_1] strategies_1] styles_1] technology_2] text_2]
 traditional_1] transformed_1] trends_1] visual_2]

C. AWL Families list

AWL families: [48:51:63]

access_1] adult_1] analyse_1] area_1] author_1] challenge_1] chart_1]
 communicate_4] component_1] compute_3] consequent_1] consult_1]
 create_1] culture_2] despite_1] distinct_1] economy_1] element_1]
 formula_1] globe_4] identify_1] institute_1] interact_1] involve_1]
 issue_1] method_1] minor_1] monitor_1] motivate_1] outcome_1]
 psychology_1] publish_1] range_2] rely_1] research_1] secure_1]
 seek_2] site_2] source_1] specific_1] strategy_1] style_1]
 technology_2] text_2] tradition_1] transform_1] trend_1] visual_2]

AWL Fr non-cognate families: [families 5 : tokens 7] involve_1]
 outcome_1] range_2] seek_2] trend_1]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	221 (66.97)	265 (68.30)	740 (82.59)	82.59
K-2 Words :	55 (16.67)	63 (16.24)	89 (9.93)	92.52
K-3 Words :	41 (12.42)	42 (10.82)	50 (5.58)	98.10
K-4 Words :	6 (1.82)	6 (1.55)	6 (0.67)	98.77
K-5 Words :	2 (0.61)	3 (0.77)	3 (0.33)	99.10
K-6 Words :				
K-7 Words :	1 (0.30)	1 (0.26)	1 (0.11)	99.21
K-8 Words :	1 (0.30)	1 (0.26)	1 (0.11)	99.32
K-9 Words :	2 (0.61)	2 (0.52)	2 (0.22)	99.54
K-10 Words :				
K-11 Words :				
K-12 Words :				
K-13 Words :				

K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :	1 (0.30)	1 (0.26)	1 (0.11)	99.65
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	3 (0.77)	3 (0.33)	99.98
Total (unrounded)	330+?	388 (100)	896 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	896
Different words (types):	388
Type-token ratio:	0.43
Tokens per type:	2.31

Pertaining to onlist only

Tokens:	893
Types:	385
Families:	330
Tokens per Family :	2.71
Types per Family :	1.17

Current profile (token %)	
K-1 (82.59)	82.59
K-2 (9.93)	92.52
K-3 (5.58)	98.10

K-4 (0.67)	98.77
K-5 (0.33)	99.10
K-7 (0.11)	99.21
K-8 (0.11)	99.32
K-9 (0.22)	99.54
K-18 (0.11)	99.65
OFF (0.33)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 191 : types 228 : tokens 685]

a [11] able [2] about [4] above [1] across [1] after [1] all [2] allow [1] already [1] also [7] always [1] an [3] and [36] answers [1] any [1] anything [1] are [10] areas [1] around [1] as [11] asked [1] at [5] backs [1] basic [1] be [11] become [2] becomes [2] been [1] being [3] below [1] better [1] between [3] biggest [1] books [1] both [1] bottom [1] brazil [2] bringing [1] but [3] by [7] central [1] change [2] changing [3] children [3] choose [1] class [2] come [3] coming [1] computer [1] computers [2] concerned [1] considering [1] countries [1] cross [1] dark [1] decide [1] difference [1] different [3] difficult [1] do [1] done [1] educational [2] ever [2] exactly [1] expect [2] expected [1] experience [1] fact [1] fall [1] fast [2] fewer [1] find [1] first [1] follow [1] football [1] for [7] friends [1] from [4] fuller [1] fun [1] games [1] general [2] gets [1] getting [1] give [1] go [2] greater [1] growing [2] has [3] have [4] having [1] hearing [1] highly [1] his [2] how [3] important [2] in [19] internet [1] involved [1] is [18] issues [1] it [7] its [1] join [1] joining [1] just [1] keep [1] kind [3] lastly [1] learn [7] learner [2] learners [14] learning [13] leave [1] less [3] life [1] like [1] looking [1] losing [1] made [1] mainly [1] make [1] many [1] may [5] means [1] moment [1] more [10] most [2] much [1] must [1] names [1] need [1] needed [1] needs [2] new [1] next [1] not [2] now [1] number [4] of [36] often [2] older [2] on [2] one [4] or [4] out [1] over [1] page [1] pain [1] parents [1] part [1] particular [1] past [1] people [6] perhaps [1] phones [1] picture [4] place [1] players [1] position [1] programs [1] questions [1] rather [2] read [1] reason [1] reasons [1] relations [1] remembering [1] room [1] said [1] same [1] sammay [1] school [1] security [1] service [1] set [1] side [1] so [3] some [1] something [1] sounding [1] speak [1] speaker [1] speakers [1] special [2] specialised [1] spoken [1] students [1] such [1] suits [1] support [3] supporting [1] sure [1] take [1] taught [1] teachers [3] teaching [1] televisions [1] tell [1] than [4] that [5] the [51] their [4] then [1] these [2] they [3] thing [1] things [2] this [2] those [1] though [1] through [1] to [22] together [1] train [1] trusted [1] turning [1] up [2] us [1] use [1]

used [1] usually [1] views [2] was [2] way [1] ways [1] we [2] web [2] well [2] what [3] when [3] where [2] whether [1] which [6] who [2] why [3] wider [1] will [16] with [4] world [5] worldly [1] worried [1] worth [1] writer [1] written [1] wrote [1] you [5] young [6] younger [3]

BNC-COCA-2,000 types: [fams 55 : types 62 : tokens 90]

accessible [1] activity [1] adults [1] applied [1] article [2] brand [1] century [1] challenges [1] coping [1] council [1] create [1] cultural [2] demand [1] demanding [1] determination [1] developing [1] discussion [1] economies [1] exchange [1] february [1] foreign [1] future [12] identity [1] increasing [1] increasingly [1] instrumental [1] introducing [1] knowledge [2] language [3] languages [3] lessons [1] likely [1] material [1] materials [1] mates [1] minority [1] mixes [1] native [2] non [1] pleasure [1] populations [1] private [1] progress [1] provide [2] provided [1] providers [2] range [2] rely [1] remain [1] researcher [1] seeking [1] sites [2] skills [2] social [1] sought [1] specific [1] struggling [1] styles [1] technologies [1] technology [2] teenagers [2] therefore [1] traditional [1] typical [1]

BNC-COCA-3,000 types: [fams 41 : types 42 : tokens 50]

analysis [1] author [1] broadcaster [1] chart [1] communication [4] component [1] consequences [1] consultant [1] content [2] curriculum [1] curve [1] despite [1] dialogues [1] distinction [1] effectively [1] element [1] formula [1] global [3] globalisation [1] hint [1] independence [1] institutions [1] interactive [1] manufacturers [1] methods [1] mobile [1] monitor [1] motivation [1] outcomes [1] psychological [1] published [1] reflect [1] reform [1] reinvented [1] rewarding [1] sources [1] strategies [1] text [2] transformed [1] trends [1] universalisation [1] visual [2]

BNC-COCA-4,000 types: [fams 6 : types 6 : tokens 6]

audio [1] blurred [1] cognitive [1] discouraging [1] hardware [1] paradox [1]

BNC-COCA-5,000 types: [fams 2 : types 3 : tokens 3]

convergence [1] converging [1] dispositions [1]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams 1 : types 1 : tokens 1]

transparency [1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

[linguist_1](#)

BNC-COCA-9,000 types: [fams 2 : types 2 : tokens 2]

[envisioning_1](#) [slog_1](#)

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams 1 : types 1 : tokens 1]

[podcasts_1](#)

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 3]

[downloadable_1](#) [mpnumber_1](#) [pathway_1](#)

B. Families list

BNC-COCA-1,000 Families: [fams 191 : types 228 : tokens 685]

[a_14](#) [able_2](#) [about_4](#) [above_1](#) [across_1](#) [after_1](#) [all_2](#) [allow_1](#) [already_1](#) [also_7](#) [always_1](#) [and_36](#) [answer_1](#) [any_2](#) [area_1](#) [around_1](#) [as_11](#) [ask_1](#) [at_5](#) [back_1](#) [basic_1](#) [be_45](#) [become_4](#) [below_1](#) [better_1](#) [between_3](#) [big_1](#) [book_1](#) [both_1](#) [bottom_1](#) [bring_1](#) [but_3](#) [by_7](#) [centre_1](#) [change_5](#) [child_3](#) [choose_1](#) [class_2](#) [come_4](#) [computer_3](#) [concern_1](#) [consider_1](#) [country_1](#) [cross_1](#) [dark_1](#) [decide_1](#) [difference_1](#) [different_3](#) [difficult_1](#) [do_2](#) [educate_2](#) [end_of_list_1](#) [ever_2](#) [exact_1](#) [expect_3](#) [experience_1](#) [fact_1](#) [fall_1](#) [fast_2](#) [few_1](#) [find_1](#) [first_1](#) [follow_1](#) [football_1](#) [for_7](#) [friend_1](#) [from_4](#) [full_1](#) [fun_1](#) [game_1](#) [general_2](#) [get_2](#) [give_1](#) [go_2](#) [great_1](#) [grow_2](#) [have_8](#) [he_2](#) [hear_1](#) [high_1](#) [how_3](#) [important_2](#) [in_19](#) [internet_1](#) [involve_1](#) [issue_1](#) [it_8](#) [join_2](#) [just_1](#) [keep_1](#) [kind_3](#) [last_1](#) [learn_36](#) [leave_1](#) [less_3](#) [life_1](#) [like_1](#) [look_1](#) [lose_1](#) [main_1](#) [make_2](#) [many_1](#) [may_5](#) [mean_1](#) [moment_1](#) [more_10](#) [most_2](#) [much_1](#) [must_1](#) [name_1](#) [need_4](#) [new_1](#) [next_1](#) [not_2](#) [now_1](#) [number_4](#) [of_36](#) [often_2](#) [old_2](#) [on_2](#) [one_4](#) [or_4](#) [out_1](#) [over_1](#) [page_1](#) [pain_1](#) [parent_1](#) [part_1](#) [particular_1](#) [past_1](#) [people_6](#) [perhaps_1](#) [picture_4](#) [place_1](#) [play_1](#) [position_1](#) [programme_1](#) [question_1](#) [rather_2](#) [read_1](#) [reason_2](#) [relate_1](#) [remember_1](#) [room_1](#) [same_1](#) [say_1](#) [school_1](#) [secure_1](#) [service_1](#) [set_1](#) [side_1](#) [so_3](#) [some_2](#) [sound_1](#) [speak_4](#) [special_3](#) [student_1](#) [such_1](#) [suit_1](#) [support_4](#) [sure_1](#) [take_1](#) [teach_5](#) [telephone_1](#) [television_1](#) [tell_1](#) [than_4](#) [that_6](#) [the_51](#) [then_1](#) [they_7](#) [thing_3](#) [this_4](#) [though_1](#) [through_1](#) [to_22](#) [together_1](#) [train_1](#) [trust_1](#) [turn_1](#) [up_2](#) [use_2](#) [usual_1](#) [view_2](#) [way_2](#) [we_3](#) [web_2](#) [well_2](#) [what_3](#) [when_3](#) [where_2](#) [whether_1](#) [which_6](#) [who_2](#) [why_3](#) [wide_1](#) [will_16](#) [with_4](#) [world_6](#) [worry_1](#) [worth_1](#) [write_3](#) [you_5](#) [young_9](#)

BNC-COCA-2,000 Families: [fams 55 : types 62 : tokens 90]

[access_1](#) [active_1](#) [adult_1](#) [apply_1](#) [article_2](#) [brand_1](#) [century_1](#) [challenge_1](#) [cope_1](#) [council_1](#) [create_1](#) [culture_2](#) [demand_2](#) [determine_1](#) [develop_1](#) [discuss_1](#) [economy_1](#) [exchange_1](#) [february_1](#) [foreign_1](#) [future_12](#) [identify_1](#) [increase_2](#) [instrument_1](#) [introduce_1](#) [knowledge_2](#) [language_6](#) [lesson_1](#) [likely_1](#) [mate_1](#) [material_2](#) [minor_1](#) [mix_1](#) [native_2](#) [non_1](#) [pleasure_1](#) [population_1](#) [private_1](#) [progress_1](#) [provide_5](#) [range_2](#) [rely_1](#) [remain_1](#) [research_1](#) [seek_2](#) [site_2](#) [skill_2](#) [social_1](#) [specific_1](#) [struggle_1](#) [style_1](#) [technology_3](#) [teenage_2](#) [therefore_1](#) [tradition_1](#) [typical_1](#)

BNC-COCA-3,000 Families: [fams 41 : types 42 : tokens 50]

analyse_[1] author_[1] broadcast_[1] chart_[1] communicate_[4]
component_[1] consequence_[1] consult_[1] content_[2] curriculum_[1]
curve_[1] despite_[1] dialogue_[1] distinct_[1] effective_[1] element_[1]
formula_[1] global_[4] hint_[1] independence_[1] institution_[1] interact_[1]
invent_[1] manufacture_[1] method_[1] mobile_[1] monitor_[1] motive_[1]
outcome_[1] psychology_[1] publish_[1] reflect_[1] reform_[1] reward_[1]
source_[1] strategy_[1] text_[2] transform_[1] trend_[1] universe_[1]
visual_[2]

BNC-COCA-4,000 Families: [fams 6 : types 6 : tokens 6]

audio_[1] blur_[1] cognitive_[1] discourage_[1] hardware_[1] paradox_[1]

BNC-COCA-5,000 Families: [fams 2 : types 3 : tokens 3]

converge_[2] disposition_[1]

BNC-COCA-6,000 Families: [fams : types : tokens]

BNC-COCA-7,000 Families: [fams 1 : types 1 : tokens 1]

transparency_[1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

linguist_[1]

BNC-COCA-9,000 Families: [fams 2 : types 2 : tokens 2]

envision_[1] slog_[1]

BNC-COCA-10,000 Families: [fams : types : tokens]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams 1 : types 1 : tokens 1]

podcast_[1]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 3]

downloadable_[1] mpnumber_[1] pathway_[1]

7.2.3.5 *Native Americans: We Are Still Here*

7.2.3.5.1 Text only file

BEFORE YOU READ

Brainstorm in class : What do you know about the history of Native Americans and their situation today ? How many American Indian tribes can you name ?

Native Americans : We Are Still Here

The ancestors of today's Native Americans probably migrated over the Bering Strait from Asia to North America between ten thousand and fifteen thousand years ago . Prior to contact with Europeans , Native Americans lived in diverse groups across North America . Some , such as the Cheyenne and Sioux , were nomadic and roamed from one area to another , usually following seasonal patterns of animal migration . Some , such as the Iroquois and Pueblo , were farmers . Many cultivated the famous " three sisters " of corn , beans , and squash . Coastal people , such as the Tlingit , fished . Many groups , such as the Yurok of northern California , combined hunting , fishing , and gathering . While native religions differed from one another and changed over time , most appear to have included significant attention to the natural world . For example , origin stories that Native Americans tell to

explain their beginnings often feature animals , plants , the moon , the stars , mountains , valleys , and rivers .

Today's Native Americans are the descendants of these first people to live in North America . In fact , Canadians use the terms " First Peoples " and " First Nations " to describe them . In the United States , most Native Americans prefer to use their tribal name . " Native American " is probably the term Americans use most when referring to all native people . Other terms include American Indian , Indian , or Native American Indian . When discussing Native Americans in reference to other first peoples around the world , like the Aboriginal people of Australia , the Maori of New Zealand or the Sami in Scandinavia , a common term is " indigenous peoples " .

Like many indigenous peoples , Native Americans experienced conquest and colonization . The most devastating aspect of colonization for native people was disease . Diseases new to Native American people , like smallpox , significantly reduced native populations upon first contact with Europeans in the 1500s . Perhaps the most famous Native American victim of smallpox was Pocahontas , who contracted the disease while in London with her husband , the Englishman John Rolfe . In addition to germs , Europeans brought with them new trade goods , different understandings of land ownership , and Christian religions . Native Americans negotiated with Europeans , sometimes peacefully and sometimes violently , over land settlement and trade . Disease decreased native populations more than

warfare . Trade , land settlement , and new religious practices dramatically changed Native American lives by the late 1800s .

Reservations

Perhaps the most far reaching legacy of the colonization period for Native Americans is the reservation system . Treaties that ended wars between Native American tribes and the US government often created reservations , pieces of land guaranteed to specific tribes . For nomadic native people , reservations sometimes functioned as prisons in the late nineteenth century because they confined native people to a limited area and prevented them from hunting . Although treaties required the US government to provide food and education for tribes on reservations , corruption sometimes prevented supplies from reaching the people who needed them , and hunger

was common on many reservations in the late nineteenth century . Most Native Americans were forbidden from practicing their traditional religions . Many Native American children attended reservation schools or boarding schools that tried to assimilate native people by teaching exclusively in English and encouraging white American norms . The Wounded Knee Massacre of 1890 occurred in this context . At Wounded Knee , hungry and confused by the new constraints on their lives , many Sioux became followers of a new religion called the Ghost Dance Movement . The Ghost Dance disturbed US Army members . Tensions were high when shots were fired at the Pine Ridge Reservation . When the shooting ended , over one hundred Sioux and over twenty five US soldiers had died .

Wounded Knee is a common , but false , end point in Native American history . It is an understandable ending to the Plains Wars between Native Americans and the US government , but many tribes did not live on the Plains . Also , not all native people who lived on the Plains played a role in the Plains Wars . Most significantly , if all stories ended at Wounded

Knee , one would think that Native Americans disappeared , but Native Americans did not disappear .

Native Americans persevered . Surprisingly , reservations gradually became home lands and political centers for indigenous people , even those whose reservations were far from their original territory . Many of those Native Americans who had been educated in English and American government used their knowledge of American law to advocate for their tribes . Some created multi tribal organizations that advocated for Native Americans broadly . By the late twentieth century , vibrant movements on behalf of Native Americans existed across the United States .

Today , Native Americans live much like other Americans . They work in dozens of occupations . Some are teachers , lawyers , and doctors . Others are artists , musicians , and actors . They are full citizens of the US , vote in federal elections , and many serve in the US military . Visitors to the US who expect Native Americans to wear feathered headdresses and beaded buckskin are surprised to learn that most native people live in cities and use modern conveniences like electricity and indoor plumbing . Because of the disadvantages that they faced historically , Native Americans are Students wearing traditional tribal outfits at the Wellpinit Elementary/High School on the Spokane Indian Reservation , Washington State .

disproportionately represented among the poor and alcoholic in the US Native Americans work hard to address such social problems today .

Sovereignty

What makes Native Americans unique in the US is their political status . Other racial minorities in the United States like African Americans , Asian Americans , and Latinos have made the rights of full US citizenship their top political priority . In contrast , Native Americans have focused on sovereignty . Native Americans who live on reservations are a part of their tribe and their nation , both sovereign bodies . They vote in tribal elections , and their tribal government negotiates with the federal government .

Three key issues dominate contemporary tribal negotiations with the US government . The first is treaty obligations . Some tribal governments believe that the US has not maintained the treaty obligations from treaties signed in the past . Many tribal governments push for the US government to meet those obligations today through funds for education , health care , and infrastructure development . A second issue is the protection of sacred lands and artifacts . Many tribes consider parts of the landscape sacred . Not all such sacred sites lie on reservations . Tribes ask that the US government protect such sacred sites . A third issue is business expansion and environmental protection . Like all people , Native Americans balance demands on natural resources with protection of natural areas . Because tribes are sovereign , tribal governments represent their people when the federal

fiLM CORNER

government , a state government , or a corporation wants to use the natural resources on a reservation for profit .

For all Native Americans a persistent frustration has been the use of Native American figures as sports mascots and advertising symbols . As the twenty first century began , a multi tribal organization called the National Congress of American Indians asked the American

football team , the Washington Redskins , to change their name . Many have also criticized the football team ,the Kansas City Chiefs , as well as two baseball teams , the Atlanta Braves and the Cleveland Indians , for using Native American images to promote their teams . Native Americans insist that such mascots reduce native people to appearances and historical stereotypes . Mascots confuse non native people by suggesting that Native Americans died out or are imaginary creatures . Even as Native Americans are in the process of convincing national sports teams to change their mascots , they continue to struggle against stereotypes of native people common in blockbuster movies and popular television shows . They particularly object to the idea that they vanished . To remind non native people of their political presence , their modernity , and their common humanity , Native Americans often use the phrase : " We are still here . "

(Flannery Burke , Associate Professor , St Louis University June 2014 .)

Title : Bury My Heart at Wounded Knee

Category : Drama (television film based on the 1970 book of the same name by Dee Brown)
Production Year : 2007

Country : USA Languages : English Director : Yves Simoneau Runtime: 132 minutes

Main Cast : Aidan Quinn, Adam Beach, August Schellenberg, Anna Paquin

Plot : We follow Lakota chief Sitting Bull and the

Sioux doctor Charles Eastman through the eventful period from the 1870s till the Wounded Knee Massacre in 1890 . We see the consequences of the Dawes Act of 1887 , which allowed for the president to break up reservation land to give 160 acres each to individual settlers .

7.2.3.5.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

3. Hyphenated words with hyphen removed:

far-reaching, twenty-five, multi-tribal, twenty-first, non-native

4. Compound words separated:

endpoint, homelands

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Americans, Bering, Asia, America, Cheyenne, Sioux, Iroquois, Pueblo, Tlingit, Yurok, California, Canadians, Indian, American, Australia, Maori, Zealand, Sami, Scandinavia, Englishman, John, Rolfe, English, African, Americans , Asian, Americans , Latinos, Washington, Redskins, Kansas, Atlanta, Cleveland, Yves, Simoneau, Aidan, Quinn, Adam, Beach, August, Schellenberg, Anna, Paquin, Lakota, Charles, Eastman, Dawes, Flannery, Burke, St Louis, June, Europeans, Pocahontas, London, Spokane, Wellpinit, Dee, USA, Indians, Christian,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Native (Americans), (Bering) Strait, North (America), Europeans, (Yurok) of northern (California), United States, First Peoples, First Nations, Native (American Indian), New (Zealand), Wounded Knee Massacre, Ghost Dance Movement, Pine Ridge Reservation, Plains Wars, the Plains, National Congress of (American Indians), (Kansas) City Chiefs, (Atlanta) Braves, Sitting Bull, (Dawes) Act, (St Louis) University

Note: Text related to illustrations have been included in the text analysis.

7.2.3.5.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Targets Native Americans We are (9.55 kb)

Words recategorized by user as 1k items (proper nouns etc): AMERICANS, BERING, ASIA, AMERICA, CHEYENNE, SIOUX, IROQUOIS, PUEBLO, TLINGIT, YUROK, CALIFORNIA, CANADIANS, INDIAN, AMERICAN, AUSTRALIA, MAORI, ZEALAND, SAMI, SCANDINAVIA, ENGLISHMAN, JOHN, ROLFE, ENGLISH, AFRICAN, AMERICANS , ASIAN, AMERICANS , LATINOS, WASHINGTON, REDSKINS, KANSAS, ATLANTA, CLEVELAND, YVES, SIMONEAU, AIDAN, QUINN, ADAM, BEACH, AUGUST, SCHELLENBERG, ANNA, PAQUIN, LAKOTA, CHARLES, EASTMAN, DAWES, FLANNERY, BURKE, ST LOUIS, JUNE, EUROPEANS, POCAHONTAS, LONDON, SPOKANE, WELLPINIT, DEE, USA, INDIANS, CHRISTIAN (total 126 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	265	333	1006	71.35%
Function:	(496)	(35.18%)
Content:	(510)	(36.17%)
> Anglo-Sax	(195)	(13.83%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	52	60	100	7.09%
> Anglo-Sax:	(26)	(1.84%)
1k+2k	(78.44%)
AWL Words (academic):	58	65	79	5.60%
> Anglo-Sax:	(7)	(0.50%)
Off-List Words:	?	<u>131</u>	<u>225</u>	<u>15.96%</u>
	375+?	589	1410	100%
Words in text (tokens):	1410			
Different words (types):	589			
Type-token ratio:	0.42			
Tokens per type:	2.39			
Lex density (content words/total)	0.65			

Pertaining to onlist only

Tokens: 1185

Types: 458
 Families: 375
 Tokens per family: 3.16
 Types per family: 1.22
 Anglo-Sax Index: %
(A-Sax tokens + functors / onlist tokens)
 Greco-Lat/Fr-Cognate Index: (Inverse of above) %

Current profile	
%	Cumul.
71.35	71.35
7.09	78.44
5.60	84.04
15.96	100.00

A. AWL Tokens lists

AWL [58:65:79] advocate advocated area area areas aspect behalf category confined consequences constraints contact contact contemporary context contracted contrast convincing corporation created created disproportionately diverse dominate drama dramatically environmental exclusively expansion feature federal federal focused functioned funds guaranteed images individual infrastructure issue issue issues maintained migrated migration military minorities norms occupations occurred period period persistent prior priority process promote required resources resources role significant significantly significantly sites sites specific status symbols team team teams teams teams tensions traditional traditional unique

Sublist 1

area area areas context contracted created created environmental functioned individual issue issue issues occurred period period process required role significant significantly significantly specific

Sublist 2

aspect category consequences feature focused maintained resources resources sites sites traditional traditional

Sublist 3

constraints corporation disproportionately dominate exclusively funds minorities

Sublist 4

contrast occupations prior promote status

Sublist 5

contact contact expansion images symbols

Sublist 6

diverse federal federal federal migrated migration

Sublist 7

advocate advocated guaranteed priority unique

Sublist 8

contemporary drama dramatically infrastructure tensions

Sublist 9

behalf confined military norms team team teams teams teams

Sublist 10

convincing persistent

B. AWL Types list

AWL types: [58:65:79] advocate_[1] advocated_[1] area_[2] areas_[1] aspect_[1] behalf_[1] category_[1] confined_[1] consequences_[1] constraints_[1] contact_[2] contemporary_[1] context_[1] contracted_[1] contrast_[1] convincing_[1] corporation_[1] created_[2] disproportionately_[1] diverse_[1] dominate_[1] drama_[1] dramatically_[1] environmental_[1] exclusively_[1] expansion_[1] feature_[1] federal_[3] focused_[1] functioned_[1] funds_[1] guaranteed_[1] images_[1] individual_[1] infrastructure_[1] issue_[2] issues_[1] maintained_[1] migrated_[1] migration_[1] military_[1] minorities_[1] norms_[1] occupations_[1] occurred_[1] period_[2] persistent_[1] prior_[1] priority_[1] process_[1] promote_[1] required_[1] resources_[2] role_[1] significant_[1] significantly_[2] sites_[2] specific_[1] status_[1] symbols_[1] team_[2] teams_[3] tensions_[1] traditional_[2] unique_[1]

C. AWL Families list

AWL families: [58:65:79] advocate_[2] area_[3] aspect_[1] behalf_[1] category_[1] confine_[1] consequent_[1] constrain_[1] contact_[2] contemporary_[1] context_[1] contract_[1] contrast_[1] convince_[1] corporate_[1] create_[2] diverse_[1] dominate_[1] drama_[2] environment_[1] exclude_[1] expand_[1] feature_[1] federal_[3] focus_[1] function_[1] fund_[1] guarantee_[1] image_[1] individual_[1] infrastructure_[1] issue_[3] maintain_[1] migrate_[2] military_[1] minor_[1] norm_[1] occupy_[1] occur_[1] period_[2] persist_[1]

prior_[1] priority_[1] process_[1] promote_[1] proportion_[1] require_[1]
 resource_[2] role_[1] significant_[3] site_[2] specific_[1] status_[1]
 symbol_[1] team_[5] tense_[1] tradition_[2] unique_[1]

AWL Fr non-cognate families: [families 3 : tokens 7] behalf_[1] feature_[1]
 team_[5]

2. VP-Compleat

Frequency framework is «BNC-COCA» - Input Mode is WINDOW - smaller texts but richer information (integral, edit, proper, cognates, extraction, barchart)

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	287 (57.98)	349 (59.05)	1008 (70.74)	70.74
K-2 Words :	109 (22.02)	123 (20.81)	231 (16.21)	86.95
K-3 Words :	56 (11.31)	66 (11.17)	98 (6.88)	93.83
K-4 Words :	21 (4.24)	21 (3.55)	27 (1.89)	95.72
K-5 Words :	7 (1.41)	7 (1.18)	11 (0.77)	96.49
K-6 Words :	4 (0.81)	4 (0.68)	4 (0.28)	96.77
K-7 Words :	6 (1.21)	6 (1.02)	8 (0.56)	97.33
K-8 Words :	1 (0.20)	1 (0.17)	1 (0.07)	97.40
K-9 Words :	1 (0.20)	1 (0.17)	2 (0.14)	97.54
K-10 Words :	2 (0.40)	2 (0.34)	5 (0.35)	97.89
K-11 Words :				
K-12 Words :				
K-13 Words :				
K-14 Words :				
K-15 Words :	1 (0.20)	1 (0.17)	1 (0.07)	97.96
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				

K-24 Words :

K-25 Words :

Off-List:	??	2 (0.34)	2 (0.14)	98.10
Total (unrounded)	495+?	591 (100)	1425 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1425
Different words (types):	591
Type-token ratio:	0.41
Tokens per type:	2.41

Pertaining to onlist only

Tokens:	1423
Types:	589
Families:	495
Tokens per Family :	2.87
Types per Family :	1.19

Current profile (token %)	
K-1 (70.74)	70.74
K-2 (16.21)	86.95
K-3 (6.88)	93.83
K-4 (1.89)	95.72
K-5 (0.77)	96.49
K-6 (0.28)	96.77
K-7 (0.56)	97.33
K-8 (0.07)	97.40
K-9 (0.14)	97.54
K-10 (0.35)	97.89
K-15 (0.07)	97.96
OFF (0.14)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 215 : types 263 : tokens 912]

a [13] about [1] across [2] act [1] actors [1] addition [1] address [1]
advertising [1] against [1] ago [1] all [6] allowed [1] also [2] although [1]
among [1] an [1] and [44] animal [1] animals [1] another [2] appear [1]
appearances [1] are [12] area [2] areas [1] around [1] artists [1] as [10]
ask [1] asked [1] at [5] based [1] beach [1] became [2] because [3]
been [2] before [1] began [1] beginnings [1] believe [1] between [3]
bodies [1] book [1] both [1] break [1] brought [1] brown [1] business [1]
but [3] by [6] called [2] can [1] care [1] centers [1] change [2]
changed [2] children [1] christian [1] cities [1] city [1] class [1]
consider [1] continue [1] corner [1] country [1] dance [2] did [2] died [2]
different [1] do [1] doctor [1] doctors [1] each [1] educated [1]
education [2] end [1] ended [3] ending [1] even [2] expect [1]
experienced [1] explain [1] faced [1] fact [1] far [2] farmers [1] fifteen [1]
figures [1] film [2] fired [1] first [7] fished [1] fishing [1] five [1] follow [1]
followers [1] following [1] food [1] football [2] for [14] from [9] full [2]
give [1] government [11] governments [3] groups [2] had [2] hard [1]
has [2] have [4] health [1] heart [1] her [1] here [2] high [2] historical [1]
historically [1] history [2] home [1] how [1] humanity [1] hundred [1]
hunger [1] hungry [1] hunting [2] husband [1] idea [1] if [1] imaginary [1]
in [30] indoor [1] is [10] issue [2] issues [4] it [1] key [1] know [1]
land [5] lands [2] late [4] law [1] learn [1] lie [1] like [7] live [5] lived [2]
lives [2] made [1] main [1] makes [1] many [13] meet [1] members [1]
minutes [1] more [1] most [9] mountains [1] movement [1]
movements [1] movies [1] much [1] musicians [1] my [1] name [4]
nation [1] national [2] nations [1] natural [4] needed [1] new [6]
nineteenth [2] north [3] not [5] number [8] numbers [3] of [34] often [3]
on [13] one [4] or [5] other [4] others [1] out [1] over [5] ownership [1]
part [1] particularly [1] parts [1] past [1] people [19] peoples [4]
perhaps [2] pieces [1] plants [1] played [1] point [1] poor [1] prisons [1]
probably [2] problems [1] protect [1] protection [3] push [1] reaching [2]
read [1] rights [1] rivers [1] s [15] same [1] school [1] schools [2]
second [1] see [1] serve [1] settlement [2] settlers [1] shooting [1]
shots [1] shows [1] signed [1] sisters [1] sitting [1] situation [1] some [5]
sometimes [4] sports [2] st [1] stars [1] state [2] still [2] stories [2]
students [1] such [8] suggesting [1] surprised [1] surprisingly [1]
system [1] teachers [1] teaching [1] team [2] teams [3] television [2]
tell [1] ten [1] term [2] terms [2] than [1] that [12] the [90] their [20]
them [4] these [1] they [8] think [1] third [1] this [1] those [3]
thousand [2] three [2] through [2] till [1] time [1] to [35] today [6] top [1]
tried [1] twentieth [1] twenty [2] two [1] understandable [1]
understandings [1] up [1] upon [1] use [7] used [1] using [1] usually [1]
visitors [1] wants [1] wars [3] was [3] we [4] wear [1] wearing [1] well [1]
were [6] what [2] when [5] which [1] while [2] white [1] who [6]
whose [1] with [8] work [2] world [2] would [1] year [1] years [1] you [3]

BNC-COCA-2,000 types: [fams 109 : types 116 : tokens 233]

alcoholic [1] army [1] associate [1] attended [1] attention [1] august [1]
balance [1] beans [1] braves [1] broadly [1] bury [1] cast [1] century [4]
chief [1] chiefs [1] citizens [1] citizenship [1] coastal [1] combined [1]
common [5] confuse [1] confused [1] contact [2] contracted [1]
convincing [1] created [2] creatures [1] demands [1] describe [1]
development [1] director [1] disappear [1] disappeared [1] discussing [1]
disease [3] diseases [1] disturbed [1] dozens [1] drama [1]
dramatically [1] elections [2] electricity [1] encouraging [1]
environmental [1] eventful [1] example [1] existed [1] famous [2]
feathered [1] feature [1] frustration [1] funds [1] gathering [1] ghost [2]
guaranteed [1] images [1] include [1] included [1] individual [1] insist [1]
june [1] knee [6] knowledge [1] languages [1] lawyers [1] limited [1]
maintained [1] military [1] minorities [1] modern [1] modernity [1]
moon [1] native [56] non [2] northern [1] object [1] occurred [1]
organization [1] organizations [1] original [1] patterns [1] peacefully [1]
period [2] pine [1] political [4] popular [1] populations [2] practices [1]
practicing [1] prefer [1] president [1] prevented [2] process [1]
production [1] provide [1] reduce [1] reduced [1] reference [1]
referring [1] remind [1] represent [1] represented [1] required [1]
reservation [6] reservations [9] role [1] seasonal [1] sites [2] social [1]
soldiers [1] specific [1] states [3] struggle [1] supplies [1] tensions [1]
title [1] trade [3] traditional [2] united [3] university [1] valleys [1]
victim [1] violently [1] vote [2] wounded [6]

BNC-COCA-3,000 types: [fams 56 : types 63 : tokens 98]

acres [1] advocate [1] advocated [1] aspect [1] category [1]
colonization [3] confined [1] congress [1] consequences [1]
constraints [1] contemporary [1] context [1] contrast [1] corporation [1]
corruption [1] criticized [1] decreased [1] descendants [1] devastating [1]
differed [1] diverse [1] dominate [1] exclusively [1] expansion [1] false [1]
federal [3] focused [1] functioned [1] goods [1] gradually [1]
landscape [1] migrated [1] migration [1] negotiated [1] negotiates [1]
negotiations [1] obligations [3] occupations [1] origin [1] persistent [1]
phrase [1] plot [1] presence [1] prior [1] priority [1] professor [1] profit [1]
promote [1] racial [1] religion [1] religions [3] religious [1] resources [2]
significant [1] significantly [2] sovereign [2] sovereignty [2] status [1]
symbols [1] territory [1] treaties [3] treaty [2] tribal [10] tribe [1] tribes [9]
unique [1]

BNC-COCA-4,000 types: [fams 21 : types 21 : tokens 27]

ancestors [1] baseball [1] beaded [1] behalf [1] boarding [1] bull [1]
conveniences [1] corn [1] cultivated [1] disadvantages [1] elementary [1]
forbidden [1] indigenous [3] infrastructure [1] legacy [1] norms [1]
outfits [1] ridge [1] sacred [4] stereotypes [2] vanished [1]

BNC-COCA-5,000 types: [fams 7 : types 7 : tokens 11]

assimilate [1] massacre [2] plains [4] plumbing [1] roamed [1] squash [1]
warfare [1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

conquest [1] disproportionately [1] strait [1] vibrant [1]

BNC-COCA-7,000 types: [fams 7 : types 7 : tokens 9]

aboriginal [1] artifacts [1] germs [1] multi [2] nomadic [2] persevered [1]
scandinavia [1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

blockbuster [1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 2]

smallpox [2]

BNC-COCA-10,000 types: [fams 3 : types 3 : tokens 6]

brainstorm [1] mascots [4] pueblo [1]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 1]

luckless [1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

headdresses [1] runtime [1]

B. Families list

BNC-COCA-1,000 Families: [fams 215 : types 263 : tokens 912]

a [14] about [1] across [2] act [2] add [1] address [1] advertise [1]
against [1] ago [1] all [6] allow [1] also [2] although [1] among [1]
and [44] animal [2] another [2] appear [2] area [3] around [1] art [1]
as [10] ask [2] at [5] base [1] be [48] beach [1] because [3] become [2]
before [1] begin [2] believe [1] between [3] body [1] book [1] both [1]
break [1] bring [1] brown [1] business [1] but [3] by [6] call [2] can [1]
care [1] centre [1] change [4] child [1] city [2] class [1] consider [1]
continue [1] corner [1] country [1] dance [2] die [2] different [1] do [3]
doctor [2] door [1] each [1] educate [3] end [5] end_of_list [1] even [2]
expect [1] experience [1] explain [1] face [1] fact [1] far [2] farm [1]
figure [1] film [2] fire [1] first [7] fish [2] five [2] follow [3] food [1]
football [2] for [14] from [9] full [2] give [1] govern [14] group [2] hard [1]
have [8] health [1] heart [1] here [2] high [2] history [4] home [1]
how [1] human [1] hundred [1] hunger [2] hunt [2] husband [1] i [1]
idea [1] if [1] imagine [1] in [30] issue [3] it [1] key [1] know [1] land [7]
late [4] law [1] learn [1] lie [1] like [7] live [9] main [1] make [2]
many [13] meet [1] member [1] minute [1] more [1] most [9]
mountain [1] move [2] movie [1] much [1] music [1] name [4] nation [4]
nature [4] need [1] new [6] nine [2] north [3] not [5] number [11] of [34]
often [3] on [13] one [4] or [5] other [5] out [1] over [5] owned [1]
part [2] particular [1] past [1] people [23] perhaps [2] piece [1] plant [1]

play [1] point [1] poor [1] prison [1] probably [2] problem [1] protect [4]
push [1] reach [2] read [1] rights [1] river [1] same [1] school [3]
second [1] see [1] serve [1] settle [3] she [1] shoot [2] show [1] sign [1]
sister [1] sit [1] situation [1] some [9] sport [2] star [1] state [2] still [2]
story [2] street [1] student [1] such [8] suggest [1] surprise [2]
system [1] teach [2] team [5] television [2] tell [1] ten [1] term [4]
than [1] that [15] the [90] they [32] think [1] this [2] thousand [2]
three [3] through [2] till [1] time [1] to [35] today [6] top [1] try [1]
twenty [3] two [1] understand [2] up [1] upon [1] use [9] usual [1]
visit [1] want [1] war [3] we [4] wear [2] well [1] what [2] when [5]
which [1] while [2] white [1] who [7] with [8] work [2] world [2] would [1]
year [2] you [3]

BNC-COCA-2,000 Families: [fams 109 : types 116 : tokens 233]

alcohol [1] army [1] associate [1] attend [1] attention [1] august [1]
balance [1] bean [1] brave [1] broad [1] bury [1] cast [1] century [4]
chief [2] citizen [2] coast [1] combine [1] common [5] confuse [2]
contact [2] contract [1] convince [1] create [2] creature [1] demand [1]
describe [1] develop [1] directed [1] disappear [2] discuss [1] disease [4]
disturb [1] dozen [1] drama [2] elect [2] electric [1] encourage [1]
environment [1] event [1] example [1] exist [1] famous [2] feather [1]
feature [1] frustrate [1] fund [1] gather [1] ghost [2] guarantee [1]
image [1] include [2] individual [1] insist [1] june [1] knee [6]
knowledge [1] language [1] lawyer [1] limit [1] maintain [1] military [1]
minor [1] modern [2] moon [1] native [56] non [2] northern [1] object [1]
occur [1] organize [2] original [1] pattern [1] peace [1] period [2] pine [1]
politics [4] popular [1] population [2] practise [2] prefer [1] president [1]
prevent [2] process [1] product [1] provide [1] reduce [2] refer [2]
remind [1] represent [2] require [1] reserve [15] role [1] season [1]
site [2] social [1] soldier [1] specific [1] states [3] struggle [1] supply [1]
tense [1] title [1] trade [3] tradition [2] unite [3] university [1] valley [1]
victim [1] violent [1] vote [2] wound [6]

BNC-COCA-3,000 Families: [fams 56 : types 63 : tokens 98]

acre [1] advocate [2] aspect [1] category [1] colony [3] confine [1]
congress [1] consequence [1] constrain [1] contemporary [1] context [1]
contrast [1] corporate [1] corrupt [1] criticise [1] decrease [1] descend [1]
devastate [1] differ [1] diverse [1] dominate [1] exclusive [1]
expansion [1] false [1] federal [3] focus [1] function [1] goods [1]
gradual [1] landscape [1] migrate [2] negotiate [3] oblige [3]
occupation [1] origin [1] persist [1] phrase [1] plot [1] presence [1]
prior [1] priority [1] professor [1] profit [1] promote [1] racial [1]

religion [4] religious [1] resource [2] significant [3] sovereign [4]
status [1] symbol [1] territory [1] treaty [5] tribe [20] unique [1]

BNC-COCA-4,000 Families: [fams 21 : types 21 : tokens 27]

ancestor [1] baseball [1] bead [1] behalf [1] boarder [1] bull [1]
convenience [1] corn [1] cultivate [1] disadvantage [1] elementary [1]
forbid [1] indigenous [3] infrastructure [1] legacy [1] norm [1] outfit [1]
ridge [1] sacred [4] stereotype [2] vanish [1]

BNC-COCA-5,000 Families: [fams 7 : types 7 : tokens 11]

assimilate [1] massacre [2] plains [4] plumbing [1] roam [1] squash [1]
warfare [1]

BNC-COCA-6,000 Families: [fams 4 : types 4 : tokens 4]

conquest [1] disproportion [1] strait [1] vibrant [1]

BNC-COCA-7,000 Families: [fams 7 : types 7 : tokens 9]

aborigine [1] artifact [1] germ [1] multi [2] nomad [2] persevere [1]
scandinavia [1]

BNC-COCA-8,000 Families: [fams 1 : types 1 : tokens 1]

blockbuster [1]

BNC-COCA-9,000 Families: [fams 1 : types 1 : tokens 2]

smallpox [2]

BNC-COCA-10,000 Families: [fams 3 : types 3 : tokens 6]

brainstorm [1] mascot [4] pueblo [1]

BNC-COCA-11,000 Families: [fams : types : tokens]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 1]

buckskin_[1]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

headresses_[1] runtime_[1]

7.2.3.6 Australia – the Island Continent

7.2.3.6.1 Text only file

Australia - the island continent

before you read do you know any famous Australians did you know that there are 1,500 species of Australian spiders ?

What else do you know about Australia ?

FAST FACTS

Population - 23.5 million

Official name - Commonwealth of Australia

Capital - Canberra

Largest cities - Sydney , Melbourne , Brisbane , Perth , Adelaide . Australia has around 10 percent of the world's biodiversity .

Of the estimated 20,000 species of vascular plants found in Australia , 16,000 are found nowhere else in the world .

Of the 378 species of mammals in Australia , more than 80 percent are unique to Australia .

Of the 869 types of Australian reptile , 773 are found nowhere else .

Australia is sometimes called " the land down under " , because a lot of things are kind of " upside down " . Australia's clock and seasons reflect the

fact that it is in the southern hemisphere and opposite from us . Many forms of animal and plant species are unique to Australia . Its geology is rich and

old , because Australia once was joined to the ancient South Polar land mass before breaking away and becoming the world's largest island .

History

Before the Europeans arrived , indigenous Aboriginal people lived undisturbed in family groups for 60,000 years . Their descendants still have a special relationship with the land and with the specific area of their origins .

As recently as 1788 , Australia was " discovered " by the British who colonised it as their own . This event is now commonly described as " the invasion " by indigenous Australians . At first , the colony was used as a penal settlement , and later , waves of settlers and migrants arrived to " tame " the land and develop a new society built on European cultural ways . The indigenous population was largely ignored or exploited for several generations . In the 1960s attitudes to Aboriginal people turned around , and in 1967 the Constitution was changed so the Aboriginal people were given better rights and were finally recognised as full Australian citizens . Australia now takes its place as a multicultural nation within world politics as part of the modern Australasian region . Population

The Australian population is small in world terms , at 23.5 million people in 2014 . It is increasing at a rate of just over 1.7 percent a year .

Its intensely diverse social mix results from historical factors , and is largely made up of Indigenous Australians , descendants of early settlers and 20 century migrants from many origins . Up until the end of World War Two , Australia was predominantly an Anglo Celtic society . At the end of World War two , there was an influx of migrants from Southern and Eastern Europe , especially Greece and Italy .The past 20 years has seen an increase in migrants from Asia and also Africa . People from Asia now make up 9 percent of the population .

Ninety percent of the population live on just 0.22 percent of the total land mass . The majority (80 percent) live within 50 kilometers of the coast to avoid the harsh and difficult living conditions of the Australian outback . Each State has only one or two major coastal cities ; elsewhere there are small communities clustered around rural centres , or living widely dispersed across the vast inland " outback " .

Languages

English is the country's official language . Australian English , based on its United Kingdom origins , is characterised by both accent and idiosyncratic word use . Australian

slang is known for its laconic and warmly humorous style . Of the hundreds of Aboriginal languages that were once used , now only a handful remain , but these are being revived and built on again by the indigenous people of some regions . The languages of migrant communities from countries across the world are frequently maintained in local , family and personal settings .

How is Australia organised ?

Australia is a federal parliamentary democracy comprising six States and three Territories . Its three levels of government , federal , state or territory and local , ensure a degree of regional autonomy . Australia's position in world affairs is largely determined in the capital , Canberra . A strong public " safety net " is maintained for social security , particularly in education ,

health and housing. But private systems also exist alongside the public in these sectors , which is said to allow for greater choice for individuals and families . Official ties remain with Britain , whose Queen is also Australia's nominal Head of State.

What makes an Aussie Australian) ?

Shared values and cultural sentiments include informality , open mindedness , integrity , response to opportunity , and an independent spirit . Home and family are central to most people's lives , and house ownership is a very common aspiration . Most Australians would say that equality , justice and fairness are essential qualities in national and community life . This is often described as " a fair go for all " , and firm loyalty to friends (mates) is considered an essential Aussie characteristic .

Education

Free public education is available to all Australian children up to age 16 (17 in some states } , but some families may instead choose to pay for private or church school education . Australian children may attend Kindergarten from age four . From 5 years , primary and secondary schooling is compulsory , then many will go on to higher education , which includes universities , vocational TAFE (Technical and Further Education } colleges and some private institutions . Most Australian students attend higher education facilities in their own states and stay with their parents , rather than moving away from home . The academic year in Australia is from February or March to November or December.

Recreation , Creativity and Sport

Familiar images of Australians at play would almost certainly include beaches and surfing , barbecues and other gatherings , either with family and friends or with others of like mind in an enthusiastic crowd . Festivals and celebrations are held in all States , and for as many reasons as you can imagine , including food and wine , arts and crafts , music , theatre , film , fashion , writing , racing (horses , dogs , camels , cars , cycles } , and many popular multicultural events . Most Australians are sports mad - major sports are cricket , the Australian Rules Football League , netball , tennis and swimming . Football refers to Australian Rules Football , not soccer . Although many people follow the competitions in Europe , soccer is not huge in Australia . Netball has the highest female participation rate in Australia and Australian Rules Football the largest male . Netball is a cross between handball and basketball - handball is not a sport widely played in Australia . Most popular Australian

sports are played outdoors , reflecting a climate that does not have harsh winters , with snow and

ice being the exception rather than the rule . Thus when Australians think of the Olympics , it is the summer Olympics that are thought of and in which Australia wins the most medals . The Winter Olympics are not so well known , as there are few Australian competitors .

Tourism

A popular tourist destination , Australia has both natural and cultural treasures to explore . From the cities to the outback , the scale , scope and range of wonders and memorable experiences are as great as the land itself .

Environment

Its harsh climate , lack of water , desert interior and intensely contrasting land formations , as well as its vulnerable ecology , make farming and resource management challenging . Major environmental challenges identified by the Global World Wild life Fund include :

Deforestation , agricultural clearing and overgrazing . In some areas of Australia there is less than 25 percent left of the native vegetation present when the European settlers first came to Australia .

Overfishing and illegal fishing

The introduction of exotic species that outcompete and cause the extinction of native species . Examples include rabbits , cane toads , feral cats , camels and foxes . Pollution such as from chemicals used in farming that enter waterways and the sea . Continued population growth along the coast line destroys farm land and leads to extinction of native species of plants and animals .

Towards a sustainable Australia Like many developed countries , Australia faces challenges related to sustainability . How can future generations be able to enjoy at least the same levels of well being as people experience today ? The Sustainable Australia Report was released by the government in 2013 . The report identifies challenges that Australia face , among them

An ageing population . This will put pressure on health costs , providing aged care pensions , and having enough aged care accommodation . Energy consumption , dependency on cars , and increasing cost of housing . Limited job opportunities , and consequently , population loss in regional areas of Australia .

Climate change . Rising average temperatures in Australia bring increased risk of bush fires , droughts , and rising sea levels .

Water supplies . How to ensure enough drinking water , as well as water for farming ?

Getting the balance right between economic growth and protecting the environment . How can industry and the environment co exist in a sustainable way ? Food production . How can Australia ensure an adequate supply of food for its population at the same time as developing a food exporting industry able to meet the demand for food in Asia ? Foreign investors are buying Australian land to secure food for their own countries and this is a

controversial issue . Some Australians worry that Australia may become more of a food producer for others rather than for itself . Trish McLaine and Julianne Cheek (2014) .

7.2.3.6.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

50 km (kilometers), UK (United Kingdom), AFL

3. Hyphenated words with hyphen removed:

Anglo-Celtic, open-mindedness, co-exist

4. Compound words separated:

Brushfires, coastline, landmass, farmland

5. Words (groups of letters) removed from the text analysis:

6. Proper nouns:

Australians, Australian, Australia, Canberra, Sydney, Melbourne, Brisbane, Perth, Adelaide, Aboriginal, British, European, Europeans, Australasian, Anglo, Celtic, Europe, English, Canberra, Britain, Aussie, February, March, November, December, Olympics, Trish, McLaine, Julianne, Cheek, Greece, Italy, Africa

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Commonwealth of (Australia), South Polar landmass, Southern and Eastern (Europe), Greece, Italy, Asia, Africa, United Kingdom, Head of State, Global World Wildlife Fund, Sustainable (Australia) Report,

Note: Text related to illustrations have been included in the text analysis.

7.2.3.6.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: Targets Australia the island (9.86 kb)

Words recategorized by user as 1k items (proper nouns etc): AUSTRALIANS, AUSTRALIAN, AUSTRALIA, CANBERRA, SYDNEY, MELBOURNE, BRISBANE, PERTH, ADELAIDE, ABORIGINAL, BRITISH, EUROPEAN, EUROPEANS, AUSTRALASIAN, ANGLO, CELTIC, EUROPE, ENGLISH, CANBERRA, BRITAIN, AUSSIE, FEBRUARY, MARCH, NOVEMBER, DECEMBER, OLYMPICS, TRISH, MCLAINE, JULIANNE, CHEEK, GREECE, ITALY, AFRICA (total 98 tokens)

	Families	Types	Tokens	Percent
K1 Words (1-1000):	331	408	1114	73.97%
Function:	(578)	(38.38%)

Content:	(536)	(35.59%)
> Anglo-Sax	(231)	(15.34%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	50	52	71	4.71%
> Anglo-Sax:	(20)	(1.33%)
1k+2k	(78.68%)
AWL Words (academic):	64	75	105	6.97%
> Anglo-Sax:	(2)	(0.13%)
Off-List Words:	?	125	216	14.34%
	445+?	660	1506	100%

Words in text (tokens):	1506
Different words (types):	660
Type-token ratio:	0.44
Tokens per type:	2.28
Lex density (content words/total)	0.62

Pertaining to onlist only

Tokens:	1290
Types:	535
Families:	445
Tokens per family:	2.90
Types per family:	1.20
Anglo-Sax Index:	%
(A-Sax tokens + functors / onlist tokens)	
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
73.97	73.97
4.71	78.68
6.97	85.65
14.34	100.00

A. AWL Tokens lists

AWL [64:75:105] academic accommodation adequate area areas areas attitudes available challenges challenges challenges challenging chemicals communities communities community comprising consequently constitution

consumption contrasting controversial creativity cultural cultural cultural
cycles diverse economic energy ensure ensure ensure environment
environment environment environmental estimated exploited exporting
facilities factors federal federal finally fund generations generations global
identified identifies ignored illegal images individuals institutions integrity
intensely intensely investors issue job maintained maintained major major
major majority migrant migrants migrants migrants migrants participation
percent percent percent percent percent percent percent percent
predominantly primary range region regional regional regions released
resource response scope sectors secure security specific style
sustainability sustainable sustainable sustainable technical unique unique

Sublist 1

area areas areas available constitution creativity economic environment
environment environment environmental estimated exporting factors
identified identifies illegal individuals issue major major major majority
percent percent percent percent percent percent percent percent response
sectors specific

Sublist 2

communities communities community consequently consumption cultural
cultural cultural finally institutions investors maintained maintained
participation primary range region regional regional regions resource secure
security

Sublist 3

ensure ensure ensure fund technical

Sublist 4

adequate attitudes contrasting cycles job

Sublist 5

academic challenges challenges challenges challenging energy facilities
generations generations images style sustainability sustainable sustainable
sustainable

Sublist 6

diverse federal federal ignored migrant migrants migrants migrants migrants
scope

Sublist 7

chemicals comprising global released unique unique

Sublist 8

exploited intensely intensely predominantly

Sublist 9

accommodation controversial

Sublist 10

integrity

B. AWL Types list

AWL types: [64:75:105] academic_[1] accommodation_[1] adequate_[1]
area_[1] areas_[2] attitudes_[1] available_[1] challenges_[3] challenging_[1]
chemicals_[1] communities_[2] community_[1] comprising_[1]
consequently_[1] constitution_[1] consumption_[1] contrasting_[1]
controversial_[1] creativity_[1] cultural_[3] cycles_[1] diverse_[1]
economic_[1] energy_[1] ensure_[3] environment_[3] environmental_[1]
estimated_[1] exploited_[1] exporting_[1] facilities_[1] factors_[1] federal_[2]
finally_[1] fund_[1] generations_[2] global_[1] identified_[1] identifies_[1]
ignored_[1] illegal_[1] images_[1] individuals_[1] institutions_[1] integrity_[1]
intensely_[2] investors_[1] issue_[1] job_[1] maintained_[2] major_[3]
majority_[1] migrant_[1] migrants_[4] participation_[1] percent_[8]
predominantly_[1] primary_[1] range_[1] region_[1] regional_[2] regions_[1]
released_[1] resource_[1] response_[1] scope_[1] sectors_[1] secure_[1]
security_[1] specific_[1] style_[1] sustainability_[1] sustainable_[3]
technical_[1] unique_[2]

C. AWL Families list

AWL families: [64:75:105]
academy_[1] accommodate_[1] adequate_[1] area_[3] attitude_[1]
available_[1] challenge_[4] chemical_[1] community_[3] comprise_[1]
consequent_[1] constitute_[1] consume_[1] contrast_[1] controversy_[1]
create_[1] culture_[3] cycle_[1] diverse_[1] economy_[1] energy_[1]
ensure_[3] environment_[4] estimate_[1] exploit_[1] export_[1] facilitate_[1]
factor_[1] federal_[2] final_[1] fund_[1] generation_[2] globe_[1] identify_[2]
ignorant_[1] image_[1] individual_[1] institute_[1] integrity_[1] intense_[2]
invest_[1] issue_[1] job_[1] legal_[1] maintain_[2] major_[4] migrate_[5]
participate_[1] percent_[8] predominant_[1] primary_[1] range_[1] region_[4]
release_[1] resource_[1] respond_[1] scope_[1] sector_[1] secure_[2]
specific_[1] style_[1] sustain_[4] technical_[1] unique_[2]

AWL Fr non-cognate families: [families 2 : tokens 2] range_[1] scope_[1]

2. VP-Compleat

WEB VP OUTPUT FOR FILE: australia the island cont. (10,066 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): australians australian australia canberra sydney melbourne brisbane perth adelaide aboriginal british european europeans australasian anglo celtic europe english canberra britain aussie february march november december olympics trish mclaine julianne cheek greece italy africa end_of_list

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't* => *will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	328 (58.26)	390 (59.00)	1092 (72.51)	72.51
K-2 Words :	112 (19.89)	131 (19.82)	193 (12.82)	85.33
K-3 Words :	71 (12.61)	75 (11.35)	89 (5.91)	91.24
K-4 Words :	17 (3.02)	17 (2.57)	22 (1.46)	92.70
K-5 Words :	17 (3.02)	18 (2.72)	22 (1.46)	94.16
K-6 Words :	5 (0.89)	5 (0.76)	7 (0.46)	94.62
K-7 Words :	3 (0.53)	3 (0.45)	3 (0.20)	94.82
K-8 Words :	6 (1.07)	6 (0.91)	6 (0.40)	95.22
K-9 Words :				
K-10 Words :	1 (0.18)	1 (0.15)	1 (0.07)	95.29
K-11 Words :	2 (0.36)	2 (0.30)	4 (0.27)	95.56
K-12 Words :				
K-13 Words :	1 (0.18)	1 (0.15)	1 (0.07)	95.63
K-14 Words :				
K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :				
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				

Off-List:	??	7 (1.06)	10 (0.66)	96.29
Total (unrounded)	563+?	661 (100)	1506 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1506
Different words (types):	661
Type-token ratio:	0.44
Tokens per type:	2.28

Pertaining to onlist only

Tokens:	1496
Types:	654
Families:	563
Tokens per Family :	2.66
Types per Family :	1.16

Current profile (token %)	
K-1 (72.51)	72.51
K-2 (12.82)	85.33
K-3 (5.91)	91.24
K-4 (1.46)	92.70
K-5 (1.46)	94.16
K-6 (0.46)	94.62
K-7 (0.20)	94.82
K-8 (0.40)	95.22
K-10 (0.07)	95.29
K-11 (0.27)	95.56
K-13 (0.07)	95.63
OFF (0.66)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 271 : types 317 : tokens 1048]

a [22] able [2] about [1] across [2] africa [1] again [1] age [2] aged [2] ageing [1] all [3] allow [1] almost [1] along [1] also [3] although [1]

among_1 an_9 and_70 animal_1 animals_1 any_1 are_20
area_1 areas_2 around_3 arrived_2 arts_1 as_21 at_7 away_2
based_1 be_1 beaches_1 because_2 become_1 becoming_1
before_3 being_3 better_1 between_2 both_2 breaking_1 bring_1
built_2 bush_1 but_3 buying_1 by_6 called_1 came_1 can_4
care_2 cars_2 cats_1 cause_1 central_1 centres_1 certainly_1
change_1 changed_1 children_2 choice_1 choose_1 church_1
cities_3 clearing_1 clock_1 co_1 colleges_1 considered_1
continued_1 cost_1 costs_1 countries_3 country_1 cross_1
degree_1 did_1 difficult_1 discovered_1 do_2 does_1 dogs_1
down_2 drinking_1 each_1 early_1 education_7 either_1 else_3
elsewhere_1 end_2 enjoy_1 enough_2 enter_1 especially_1
experience_1 experiences_1 face_1 faces_1 fact_1 facts_1 fair_1
fairness_1 families_2 family_4 farm_1 farming_3 fast_1 few_1
film_1 finally_1 fires_1 first_2 fishing_1 follow_1 food_7 football_4
for_15 forms_1 found_3 four_1 free_1 friends_2 from_13 full_1
further_1 getting_1 given_1 go_2 government_2 great_1 greater_1
groups_1 growth_2 handful_1 has_5 have_2 having_1 head_1
health_2 held_1 higher_2 highest_1 historical_1 history_1 home_2
horses_1 house_1 housing_2 how_5 huge_1 hundreds_1 ice_1
imagine_1 in_33 instead_1 is_28 island_2 issue_1 it_4 its_9
itself_2 job_1 joined_1 just_2 kind_1 kingdom_1 know_3
known_2 land_9 largely_3 largest_3 later_1 leads_1 least_1
left_1 less_1 levels_3 life_2 like_2 line_1 live_2 lived_1 lives_1
living_2 local_2 lot_1 mad_1 made_1 major_3 make_2 makes_1
management_1 many_7 may_3 meet_1 million_2 mind_1
mindedness_1 more_2 most_6 moving_1 music_1 name_1
nation_1 national_1 natural_1 new_1 ninety_1 not_5 now_4
number_34 numbers_1 of_51 often_1 old_1 on_7 once_2 one_1
only_2 open_1 or_8 other_1 others_2 outdoors_1 over_1 own_3
ownership_1 parents_1 part_1 particularly_1 past_1 pay_1
people_9 personal_1 place_1 plant_1 plants_2 play_1 played_2
position_1 present_1 protecting_1 public_3 put_1 queen_1
rabbits_1 racing_1 rate_2 rather_3 read_1 reasons_1 recently_1
related_1 relationship_1 report_2 rich_1 right_1 rights_1 rising_2
rule_1 rules_3 safety_1 said_1 same_2 say_1 school_1
schooling_1 sea_2 secondary_1 secure_1 security_1 seen_1
settings_1 settlement_1 settlers_3 several_1 shared_1 six_1
small_2 snow_1 so_2 some_6 sometimes_1 south_1 special_1
sport_2 sports_3 state_3 stay_1 still_1 strong_1 students_1
such_1 summer_1 swimming_1 systems_1 takes_1 terms_1
than_5 that_10 the_73 their_6 them_1 then_1 there_5 these_2
things_1 think_1 this_4 thought_1 three_2 ties_1 time_1 to_25
today_1 total_1 towards_1 turned_1 two_3 types_1 under_1
until_1 up_4 upside_1 us_1 use_1 used_3 very_1 war_2

warmly_1 was_8 water_4 waves_1 way_1 ways_1 well_4 were_3
what_2 when_2 which_3 who_1 whose_1 widely_2 wild_1 will_2
wine_1 wins_1 winter_1 winters_1 with_7 within_2 wonders_1
word_1 world_10 worry_1 would_2 writing_1 year_2 years_3
you_5

BNC-COCA-2,000 types: [fams 116 : types 133 : tokens 198]

accent_1 affairs_1 attend_2 attitudes_1 available_1 average_1
avoid_1 balance_1 capital_2 century_1 challenges_3 challenging_1
characterised_1 cheek_1 citizens_1 coast_2 coastal_1 common_1
commonly_1 communities_2 community_1 competitions_1
conditions_1 creativity_1 crowd_1 cultural_3 december_1
demand_1 described_2 desert_1 destroys_1 determined_1
develop_1 developed_1 developing_1 economic_1 energy_1
environment_3 environmental_1 equality_1 event_1 events_1
examples_1 exist_2 familiar_1 famous_1 fashion_1 february_1
female_1 firm_1 foreign_1 foxes_1 fund_1 future_1 gatherings_1
generations_2 identified_1 identifies_1 ignored_1 illegal_1
images_1 include_4 includes_1 including_1 increase_1 increased_1
increasing_2 individuals_1 industry_2 intensely_2 introduction_1
justice_1 kilometers_1 lack_1 language_1 languages_3 league_1
limited_1 loss_1 maintained_2 male_1 march_1 mass_2 mates_1
mix_1 modern_1 native_3 november_1 nowhere_2 official_3
opportunities_1 opportunity_1 opposite_1 organised_1 pensions_1
percent_8 politics_1 pollution_1 popular_3 population_10
pressure_1 private_3 producer_1 production_1 providing_1
qualities_1 range_1 recognised_1 refers_1 region_1 regional_2
regions_1 released_1 remain_2 results_1 risk_1 scale_1
seasons_1 social_2 society_2 southern_2 species_7 specific_1
spirit_1 states_4 style_1 supplies_1 supply_1 theatre_1 thus_1
tourism_1 tourist_1 undisturbed_1 united_1 universities_1 values_1

BNC-COCA-3,000 types: [fams 71 : types 75 : tokens 89]

academic_1 accommodation_1 adequate_1 agricultural_1
alongside_1 ancient_1 celebrations_1 characteristic_1 chemicals_1
climate_3 clustered_1 colonised_1 colony_1 competitors_1
comprising_1 consequently_1 constitution_1 consumption_1
continent_1 contrasting_1 controversial_1 crafts_1 cycles_1
democracy_1 descendants_2 diverse_4 eastern_1 ensure_3
enthusiastic_1 essential_2 estimated_1 exception_1 exploited_1
explore_1 exporting_1 facilities_1 factors_1 federal_2 festivals_1
formations_1 frequently_1 global_1 harsh_3 humorous_1
independent_1 institutions_1 interior_1 invasion_1 investors_1

loyalty_[1] majority_[1] net_[1] origins_[3] parliamentary_[1] participation_[1]
primary_[1] reflect_[1] reflecting_[1] resource_[1] response_[1] revived_[1]
rural_[1] scope_[1] sectors_[1] sustainability_[1] sustainable_[3]
technical_[1] temperatures_[1] tennis_[1] territories_[1] territory_[1]
treasures_[1] unique_[2] vast_[1] vulnerable_[1]

BNC-COCA-4,000 types: [fams 17 : types 17 : tokens 22]

aspiration_[1] autonomy_[1] destination_[1] dispersed_[1] exotic_[1]
geology_[1] indigenous_[5] informality_[1] integrity_[1] mammals_[1]
medals_[1] polar_[1] predominantly_[1] recreation_[1] sentiments_[1]
soccer_[2] vocational_[1]

BNC-COCA-5,000 types: [fams 17 : types 17 : tokens 22]

basketball_[1] cane_[1] commonwealth_[1] cricket_[1] dependency_[1]
droughts_[1] ecology_[1] extinction_[2] hemisphere_[1] inland_[1]
memorable_[1] migrant_[1] migrants_[4] nominal_[1] spiders_[1] surfing_[1]
tame_[1] vegetation_[1]

BNC-COCA-6,000 types: [fams 5 : types 5 : tokens 7]

barbecues_[1] camels_[2] compulsory_[1] influx_[1] multicultural_[2]

BNC-COCA-7,000 types: [fams 4 : types 4 : tokens 7]

aboriginal_[4] penal_[1] reptile_[1] toads_[1]

BNC-COCA-8,000 types: [fams 6 : types 6 : tokens 6]

biodiversity_[1] deforestation_[1] idiosyncratic_[1] kindergarten_[1]
slang_[1] vascular_[1]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

feral_[1]

BNC-COCA-11,000 types: [fams 2 : types 2 : tokens 4]

laconic_[1] outback_[3]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams 1 : types 1 : tokens 1]

overgrazing_[1]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 8 : tokens 21]

handball_[2] netball_[3] numero_[1] outcompete_[1] overfishing_[1]
tafe_[1] waterways_[1]

B. Families list

BNC-COCA-1,000 Families: [fams 271 : types 317 : tokens 1048]

a_[31] able_[2] about_[1] across_[2] again_[1] age_[5] all_[3] allow_[1]
almost_[1] along_[1] also_[3] although_[1] among_[1] and_[70] animal_[2]
any_[1] area_[3] around_[3] arrive_[2] art_[1] as_[21] at_[7] away_[2]
base_[1] be_[63] beach_[1] because_[2] become_[2] before_[3] better_[1]
between_[2] both_[2] break_[1] bring_[1] build_[2] bush_[1] but_[3] buy_[1]
by_[6] call_[1] can_[4] car_[2] care_[2] cat_[1] cause_[1] centre_[2]
certain_[1] change_[2] child_[2] choice_[1] choose_[1] church_[1] city_[3]

clear_1] clock_1] college_1] come_1] company_1] consider_1]
 continue_1] cost_2] country_4] cross_1] degree_1] difficult_1]
 discover_1] do_4] dog_1] door_1] down_2] drink_1] each_1] early_1]
 educate_7] either_1] else_4] end_2] end_of_list_1] enjoy_1] enough_2]
 enter_1] especially_1] experience_2] face_2] fact_2] fair_2] family_6]
 farm_4] fast_1] few_1] film_1] final_1] find_3] fire_1] first_2] fish_1]
 follow_1] food_7] football_4] for_15] form_1] four_1] free_1] friend_2]
 from_13] full_1] further_1] get_1] give_1] go_2] govern_2] great_2]
 group_1] grow_2] hand_1] have_8] head_1] health_2] high_3]
 history_2] hold_1] home_2] horse_1] house_3] how_5] huge_1]
 hundred_1] ice_1] imagine_1] in_33] instead_1] island_2] issue_1]
 it_15] job_1] join_1] just_2] kind_1] king_1] know_5] land_9] large_6]
 late_1] lead_1] least_1] left_1] less_1] level_3] life_2] like_2] line_1]
 live_6] local_2] lot_1] mad_1] major_3] make_4] manage_1] many_7]
 may_3] meet_1] million_2] mind_2] more_2] most_6] move_1]
 music_1] name_1] nation_2] nature_1] new_1] nine_1] not_5] now_4]
 number_35] of_51] often_1] old_1] on_7] once_2] one_1] only_2]
 open_1] or_8] other_3] over_1] own_3] owned_1] parent_1] part_1]
 particular_1] past_1] pay_1] people_9] person_1] place_1] plant_3]
 play_3] position_1] present_1] protect_1] public_3] put_1] queen_1]
 rabbit_1] race_1] rate_2] rather_3] read_1] reason_1] recent_1]
 relate_2] report_2] rich_1] right_1] rights_1] rise_2] rule_4] safe_1]
 same_2] say_2] school_2] sea_2] second_1] secure_2] see_1] set_1]
 settle_4] several_1] share_1] six_1] small_2] snow_1] so_2] some_7]
 south_1] special_1] sport_5] state_3] stay_1] still_1] strong_1]
 student_1] such_1] summer_1] swim_1] system_1] take_1] term_1]
 than_5] that_10] the_73] then_1] there_5] they_7] thing_1] think_2]
 this_6] three_2] tie_1] time_1] to_25] today_1] total_1] toward_1]
 turn_1] two_3] type_1] under_1] until_1] up_5] use_4] very_1] war_2]
 warm_1] water_4] wave_1] way_2] we_1] well_4] what_2] when_2]
 which_3] who_2] wide_2] wild_1] will_2] win_1] wine_1] winter_2]
 with_7] within_2] wonder_1] word_1] world_10] worry_1] would_2]
 write_1] year_5] you_5]

BNC-COCA-2,000 Families: [fams 116 : types 133 : tokens 198]

accent_1] affair_1] attend_2] attitude_1] available_1] average_1]
 avoid_1] balance_1] capital_2] century_1] challenge_4] character_1]
 cheek_1] citizen_1] coast_3] common_2] community_3] competition_1]
 condition_1] create_1] crowd_1] culture_3] december_1] demand_1]
 describe_2] desert_1] destroy_1] determine_1] develop_3] disturb_1]
 economy_1] energy_1] environment_4] equal_1] event_2] example_1]
 exist_2] familiar_1] famous_1] fashion_1] february_1] female_1] firm_1]
 foreign_1] fox_1] fund_1] future_1] gather_1] generation_2] identify_2]
 ignore_1] image_1] include_6] increase_4] individual_1] industry_2]

intense_2] introduce_1] justice_1] kilometre_1] lack_1] language_4]
 league_1] legal_1] limit_1] loss_1] maintain_2] male_1] march_1]
 mass_2] mate_1] mix_1] modern_1] native_3] november_1]
 nowhere_2] official_3] opportunity_2] opposite_1] organize_1]
 pension_1] percent_8] politics_1] pollute_1] popular_3] population_10]
 pressure_1] private_3] produce_1] product_1] provide_1] quality_1]
 range_1] recognize_1] refer_1] region_4] release_1] remain_2]
 result_1] risk_1] scale_1] season_1] social_2] society_2] southern_2]
 species_7] specific_1] spirit_1] states_4] style_1] supply_2] theatre_1]
 thus_1] tour_2] unite_1] university_1] value_1]

BNC-COCA-3,000 Families: [fams 71 : types 75 : tokens 89]

academy_1] accommodate_1] adequate_1] agriculture_1] alongside_1]
 ancient_1] celebrate_1] characteristic_1] chemical_1] climate_3]
 cluster_1] colony_2] compete_1] comprise_1] consequent_1]
 constitution_1] consumption_1] continent_1] contrast_1] controversy_1]
 craft_1] cycle_1] democracy_1] descend_2] diverse_1] eastern_1]
 ensure_3] enthusiastic_1] essential_2] estimate_1] exception_1]
 exploit_1] explore_1] export_1] facility_1] factor_1] federal_2]
 festival_1] formation_1] frequent_1] global_1] harsh_3] humour_1]
 independent_1] institution_1] interior_1] invasion_1] invest_1] loyal_1]
 majority_1] net_1] origin_3] parliament_1] participate_1] primary_1]
 reflect_2] resource_1] response_1] revive_1] rural_1] scope_1]
 sector_1] sustain_4] technical_1] temperature_1] tennis_1] territory_2]
 treasure_1] unique_2] vast_1] vulnerable_1]

BNC-COCA-4,000 Families: [fams 17 : types 17 : tokens 22]

aspire_1] autonomy_1] destination_1] disperse_1] exotic_1] geology_1]
 indigenous_5] informal_1] integrity_1] mammal_1] medal_1] polar_1]
 predominant_1] recreation_1] sentiment_1] soccer_2] vocation_1]

BNC-COCA-5,000 Families: [fams 17 : types 17 : tokens 22]

basketball_1] cane_1] commonwealth_1] cricket_1] dependency_1]
 drought_1] ecology_1] extinct_2] hemisphere_1] inland_1]
 memorable_1] migrant_5] nominal_1] spider_1] surf_1] tame_1]
 vegetation_1]

BNC-COCA-6,000 Families: [fams 5 : types 5 : tokens 7]

barbecue_1] camel_2] compulsory_1] influx_1] multicultural_2]

BNC-COCA-7,000 Families: [fams 4 : types 4 : tokens 7]

aborigine_[4] penal_[1] reptile_[1] toad_[1]

BNC-COCA-8,000 Families: [fams 6 : types 6 : tokens 6]

biodiversity_[1] deforest_[1] idiosyncratic_[1] kindergarten_[1] slang_[1]
vascular_[1]

BNC-COCA-9,000 Families: [fams : types : tokens]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

feral_[1]

BNC-COCA-11,000 Families: [fams 2 : types 2 : tokens 4]

laconic_[1] outback_[3]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams 1 : types 1 : tokens 1]

overgrazed_[1]

BNC-COCA-14,000 Families: [fams : types : tokens]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams : types : tokens]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 8 : tokens 21]

handball_[2] netball_[3] numero_[1] outcompete_[1] overfishing_[1]
tafe_[1] waterways_[1]

7.2.3.7 *Ending the Era of Harmful “Indian” Mascots*

7.2.3.7.1 Text only file

Ending the Era of Harmful Indian Mascots

The invisibility of Native peoples and lack of positive images of Native cultures may not register as a problem for many Americans but it poses a significant challenge for Native youth who want to maintain a foundation in their culture and language The Washington team’s brand a name derived from historical terms for hunting native peoples is a central component to this challenge .

National Congress of American Indians President Brian Cladoosby April 2014 Washington Post opposite the editorial page NCAI’s Long Standing Opposition to Harmful Indian Sports Mascots

As the nation’s oldest largest and most representative American Indian and Alaska Native advocacy organization National Congress of American Indians has long held a clear position against derogatory and harmful stereotypes of Native people including sports mascots in media and popular culture In 1968 NCAI launched a campaign to address stereotypes of Native people in popular culture and media as well as in sports . Since this effort began there has been a great deal of progress made and support to end the era of harmful Indian mascots in sports

NCAI’s position is clear long standing and deeply rooted in our seventy years as a leading voice for Indian Country we advocate for and protect the civil rights Asocial justice and racial equity of all Native people in all parts of American society

About Indian Sports Mascots and Harm

Born in an era when racism and bigotry were accepted by the dominant culture Indian sports brands have grown to become multi million dollar franchises

The intolerance and harm promoted by these Indian sports mascots logos or symbols have very real consequences for Native peoples Specifically rather than honoring Native peoples these caricatures and stereotypes are harmful perpetuate negative stereotypes of America’s first peoples and contribute to a disregard for the personhood of Native peoples

As documented in a comprehensive review of decades of social science research derogatory Indian sports mascots have serious psychological social and cultural consequences for Native Americans especially Native youth Of today’s American Indian and Alaska Native population those under the age of 18 make up 32 percent and Native youth under the age of 24 represent nearly half or 42 percent of the entire Native population

Most concerning in considering negative stereotypes of Native people are the alarmingly high rates of hate crimes against Native people According to Department of Justice analysis American Indians are more likely than people of other races to experience violence at the hands of someone of a different race

These factors together indicate a very real need to take immediate action in a number of areas including the removal of harmful images as well as the education of the general public to diffuse additional hateful activity against Native peoples

Wide spread Support in Indian Country and Beyond

Over the last fifty years a ground swell of support has mounted to end the era of racist and harmful Indian mascots in sports and popular culture Today that support is stronger than ever

Hundreds of tribal nations national and regional tribal organizations civil rights organizations school boards sports teams sports and media personalities and individuals have called for the end to harmful Indian mascots

Rooted in the civil rights movement the quest for racial equality among American Indian and Alaska Native people began well before NCAI established a campaign in 1968 to bring an end to negative and harmful stereotypes in the media and popular culture including in sports

As a result there has been significant progress at the professional collegiate and high school levels to change once accepted race based marketing practices

The Origins of the NFL's Washington Football Team Name and Culture A Legacy of Racism

The NFL's Washington football team name Redskins is a dictionary defined racial slur The slur's origin is rooted in government bounty announcements calling for the bloody scalps of Native Americans in the 1800s From the early 1900s up until today the term has been carried on as a racial slur in popular culture For much of the 20 century the term was used interchangeably in movies and books with the word savage to portray a misleading and denigrating image of the Native American

This derogatory term was selected by team owner George Preston Marshall for use by the team in 1932 at a time when Native people were continuing to experience government and social policies to terminate tribes assimilate Native people and erase Native human and civil rights In 1932 the federal Civilization Regulations were still in place confining Native people to reservations banning all Native dances and ceremonies confiscating Native cultural property and outlawing much of what was traditional in Native life

Marshall's reputation as a segregationist and racist was only just beginning to make a mark on society and sports In 1933 Marshall was the self appointed leader amongst NFL owners to institute what would become a 13 year league wide ban on African American players from the National Football League The Washington football team did not integrate until 30 years later when Marshall was forced to do so While the team has moved on from Marshall's segregationist policies it has refused to close the chapter on Marshall's ugly use of race based marketing at the expense of Native people and communities

At the local community level 28 high schools in 18 states that have dropped the word as their mascot's name in the last 25 years

Contrary to calls for name changes by tribal nations Native peoples former players civil rights organizations media outlets and a sea change at the youth amateur collegiate and professional sports levels the Washington football team has opted to retain its harmful Indian brand Rather than truly honoring Native peoples the organization has carried on its legacy of racism and stubbornly holds on to its ugly past

Nearly 50 Years of Calling for a Change to the NFL's Washington Football Team Name Proud to Be Video and Photo Campaign

Along with the success of the Proud to Be viral video NCAI launched a photo campaign . Join the effort and submit your photo

Social Media #NotYourMascot

Started as a trend via a Twitter Storm during the 2014 NFL Superbowl the hashtag #NotYourMascot continues to illustrate Native and non Native opposition to harmful Indian Mascots Twitter Updates from Change the Mascot @ChangetheMascot

MYTH NCAI or an NCAI President gave the football team their current mascot logo

Historian Michael Richman reports in his book The Redskins Encyclopedia based on a 2002 Washington Post interview that In the early 1970s Walter Blackie Wetzel president of the National Congress of American Indians and chairman of the Blackfoot tribe urged the Redskins to replace the logo on their helmets with the head of an Indian chief

FACT Wetzel was not President of NCAI at the time he took these reported actions and these actions were not taken on behalf of NCAI's members

Wetzel served honorably as President of NCAI from 1960 to 1964 however he was not President of NCAI when he reportedly contacted the football team In 1965 the team changed the logo from an Indian mascot to a spear and in 1970 to an In 1972 the team's logo was reverted to a newer version of the original Indian mascot logo dating back to the original Boston Braves logo

NCAI and Native rights advocates have been working for nearly fifty years to change the name of the NFL's Washington team

In 1972 following the launch of the organization's campaign against Indian stereotypes representatives of NCAI the American Indian Press Association the American Indian Movement and others reached out directly to the team owner to request that the franchise change its name Since that moment in time there have been substantial efforts to call for the name change

In 1993 NCAI membership passed a resolution against the team name Resolution in Support of the Petition for Cancellation of the Registered Service Marks of the Washington Redskins AKA Pro Football Incorporated In 1999 and 2014 the U.S. Patent Office ruled that the word is disparaging to Native Americans and therefore not entitled to taxpayer financed copyright protections In 2009 NCAI filed an amicus brief along with four tribal governments (Cherokee Nation of Oklahoma Comanche Nation of Oklahoma Oneida Indian Tribe of Wisconsin and Seminole Nation of Oklahoma all federally recognized Indian tribes that have adopted resolutions condemning the use of Indian names and mascots by sports teams as well

as over 20 national Indian organizations requesting that the US Supreme Court hear an appeal to the lower court rulings and uphold the Patent Office's decision

In recent years NCAI has continued to educate the public about the issue as a new and successful legal challenge to the team name by Native youth Blackhorse Pro Football Incorporated has brought heightened attention to the issue While Native opposition to the name has not waivered public concern about the Washington football team's name has grown Indian Country and NCAI has continued to educate the public and advocate for a name change along with Native and non Native allies through the campaign Change The Mascot

7.2.1.7.2 Text changes

1. All glossary terms have been removed from the text analysis because these will be discussed separately.

2. Contractions that are written out:

NCAI (National Congress of American Indians), Op-ed (opposite the editorial page), NFL (national football league), Inc. (incorporated)

3. Hyphenated words with hyphen removed:

Anti-Defamation, multi-million, non-Native, two-thirds, 13-year, league-wide , African-American, race-based, taxpayer-financed, Pro-Football

4. Compound words separated:

Longstanding, widespread, download

5. Words (groups of letters) removed from the text analysis:

(R-)word, 20th, Mr., et al v., 's

6. Proper nouns:

Indian, Indians, Americans, Washington, Brian, Cladoosby, April, ChangeTheMascot.org, Redskin, NCAI, Alaska, American, #NotYourMascot, Twitter, NFL, @ChangetheMascot, NCAA, Marshall, Michael, Richman, Walter, Blackie, Wetzal, Boston, Cherokee, Oklahoma, Comanche, Oneida, Wisconsin, Seminole, Redskins, Blackhorse, George, Preston, African, Blackfoot,

Take note: The words outside of brackets have not been placed on the list of proper nouns.

Native peoples, Native culture, President (Brian Cladoosby), (Washington) Post, Change the Mascot, (Alaska) Native, (Indian) Country, Department of Justice, (Twitter) Storm, National Collegiate Athletic Association (NCAA), Civilization Regulations, (Boston) Braves, (Cherokee) Nation of (Oklahoma) , (Comanche) Nation of (Oklahoma) , (Oneida) Indian Tribe of (Wisconsin) , and (Seminole) Nation of (Oklahoma), US Supreme Court

Note: Text related to illustrations have been included in the text analysis.

7.2.1.7.3 Text analysis

1. VP-Classic

WEB VP OUTPUT FOR FILE: targets - Indian mascots (9.24 kb)

Words recategorized by user as 1k items (proper nouns etc): INDIAN, INDIANS, AMERICANS, WASHINGTON, BRIAN, CLADOOSBY, APRIL, CHANGETHEMASCOT.ORG, REDSKIN, NCAI, ALASKA, AMERICAN, #NOTYOURMASCOT, TWITTER, NFL, @CHANGETHEMASCOT, NCAA, MARSHALL, MICHAEL, RICHMAN, WALTER, BLACKIE, WETZEL, BOSTON, U.S., CHEROKEE, OKLAHOMA , COMANCHE, ONEIDA, WISCONSIN, SEMINOLE, REDSKINS, BLACKHORSE, GEORGE, PRESTON, AFRICAN, BLACKFOOT (total 121 tokens)

	<u>Families</u>	<u>Types</u>	<u>Tokens</u>	<u>Percent</u>
K1 Words (1-1000):	231	287	1012	67.38%
Function:	(556)	(37.02%)
Content:	(456)	(30.36%)
> Anglo-Sax	(135)	(8.99%)
=Not Greco-Lat/Fr Cog:		
K2 Words (1001-2000):	34	41	83	5.53%
> Anglo-Sax:	(33)	(2.20%)
1k+2k		(72.91%)
AWL Words (academic):	67	75	122	8.12%
> Anglo-Sax:	(23)	(1.53%)
Off-List Words:	?	138	285	18.97%
	332+?	537	1502	100%

Words in text (tokens):	1502
Different words (types):	537
Type-token ratio:	0.36
Tokens per type:	2.80
Lex density (content words/total)	0.63

Pertaining to onlist only

Tokens:	1217
Types:	403
Families:	332
Tokens per family:	3.67
Types per family:	1.21
Anglo-Sax Index:	%
<small>(A-Sax tokens + functors / onlist tokens)</small>	
Greco-Lat/Fr-Cognate Index: (Inverse of above)	%

Current profile	
%	Cumul.
67.38	67.38
5.53	72.91
8.12	81.03
18.97	100.00

A. AWL Tokens lists

AWL [67:75:122] advocacy advocate advocate advocates analysis areas behalf brief challenge challenge challenge chapter civil civil civil civil civil communities community component comprehensive confining consequences consequences contacted contrary contribute cultural cultural culture culture culture culture culture culture cultures decades defined derived documented dominant editorial established factors federal filed financed foundation illustrate image images images incorporated indicate individuals institute integrate invisibility issue legal maintain media media media media media media media negative negative negative percent percent policies policies poses positive professional professional promoted psychological regional register registered regulations removal research resolution resolution retain selected significant significant specifically submit symbols team team team team team team team team team team team team team team team team team team teams teams terminate traditional trend version via

Sublist 1

analysis areas defined derived established factors financed indicate individuals issue legal percent percent policies policies research significant significant specifically

Sublist 2

chapter communities community consequences consequences cultural cultural culture culture culture culture culture culture culture cultures institute maintain positive regional regulations selected traditional

Sublist 3

component contribute documented dominant illustrate negative negative negative negative register registered removal

civil civil civil civil civil integrate professional professional promoted resolution resolution retain

challenge challenge challenge challenge contacted image images images psychological symbols trend version

Sublist 6

brief editorial federal incorporated

Sublist 7

advocacy advocate advocate advocates comprehensive contrary decades filed foundation invisibility media media media media media media submit

Sublist 8

terminate via

Sublist 9

behalf confining team team team team team team team team team team team team team team team team team team teams teams

Sublist 10

poses

B. AWL Types list

AWL types: [67:75:122] advocacy_[1] advocate_[2] advocates_[1] analysis_[1] areas_[1] behalf_[1] brief_[1] challenge_[3] chapter_[1] civil_[5] communities_[1] community_[1] component_[1] comprehensive_[1] confining_[1] consequences_[2] contacted_[1] contrary_[1] contribute_[1] cultural_[2] culture_[8] cultures_[1] decades_[1] defined_[1] derived_[1] documented_[1] dominant_[1] editorial_[1] established_[1] factors_[1] federal_[1] filed_[1] financed_[1] foundation_[1] illustrate_[1] image_[1] images_[2] incorporated_[1] indicate_[1] individuals_[1] institute_[1] integrate_[1] invisibility_[1] issue_[1] legal_[1] maintain_[1] media_[6] negative_[3] percent_[2] policies_[2] poses_[1] positive_[1] professional_[2] promoted_[1] psychological_[1] regional_[1] register_[1] registered_[1] regulations_[1] removal_[1] research_[1] resolution_[2] retain_[1] selected_[1] significant_[2] specifically_[1] submit_[1] symbols_[1] team_[18] teams_[2] terminate_[1] traditional_[1] trend_[1] version_[1] via_[1]

C. AWL Families list

AWL families: [67:75:122] advocate_[4] analyse_[1] area_[1] behalf_[1] brief_[1] challenge_[3] chapter_[1] civil_[5] community_[2] component_[1] comprehensive_[1] confine_[1] consequent_[2] contact_[1] contrary_[1] contribute_[1] culture_[1] decade_[1] define_[1] derive_[1] document_[1] dominate_[1] edit_[1] establish_[1] factor_[1] federal_[1] file_[1] finance_[1] foundation_[1] illustrate_[1] image_[3] incorporate_[1] indicate_[1] individual_[1] institute_[1] integrate_[1] issue_[1] legal_[1] maintain_[1]

media [6] negate [3] percent [2] policy [2] pose [1] positive [1]
 professional [2] promote [1] psychology [1] region [1] register [2]
 regulate [1] remove [1] research [1] resolve [2] retain [1] select [1]
 significant [2] specific [1] submit [1] symbol [1] team [20] terminate [1]
 tradition [1] trend [1] version [1] via [1] visible [1]

AWL Fr non-cognate families: [families 4 : tokens 23] behalf [1]
 remove [1] team [20] trend [1]

2. VP-Compleat

WEB VP OUTPUT FOR FILE: ending mascot use (9,501 chars)

User Re-Cats + Mid-Sentence Capped Offlist Words => 1k: (types): indian indians americans washington brian cladoosby april changethemascot.org redskin ncai alaska american #notyourmascot twitter nfl @changethemascot ncaa marshall michael richman walter blackie wetzel boston cherokee oklahoma comanche oneida wisconsin seminole redskins blackhorse george preston african blackfoot end_of_list

Text Pre-Processing Notes: In the output text, punctuation is eliminated; all figures (1, 20, etc) are replaced by the word *number*; contractions are replaced by constituent words (*won't => will not*); type-token ratio is calculated using these modified constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns as well as program decision to class numbers as 1k although not contained in 1k list); single letters are eliminated as words except for 'a' and 'l.'

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	234 (54.29)	294 (53.85)	1068 (70.50)	70.50
K-2 Words :	87 (20.19)	100 (18.32)	198 (13.07)	83.57
K-3 Words :	61 (14.15)	72 (13.19)	112 (7.39)	90.96
K-4 Words :	17 (3.94)	18 (3.30)	26 (1.72)	92.68
K-5 Words :	12 (2.78)	13 (2.38)	18 (1.19)	93.87
K-6 Words :	5 (1.16)	5 (0.92)	5 (0.33)	94.20
K-7 Words :	5 (1.16)	5 (0.92)	5 (0.33)	94.53
K-8 Words :	2 (0.46)	2 (0.37)	4 (0.26)	94.79
K-9 Words :	3 (0.70)	3 (0.55)	4 (0.26)	95.05
K-10 Words :	1 (0.23)	2 (0.37)	17 (1.12)	96.17
K-11 Words :	1 (0.23)	1 (0.18)	3 (0.20)	96.37
K-12 Words :				
K-13 Words :				
K-14 Words :	1 (0.23)	1 (0.18)	1 (0.07)	96.44

K-15 Words :				
K-16 Words :				
K-17 Words :				
K-18 Words :	2 (0.46)	2 (0.37)	2 (0.13)	96.57
K-19 Words :				
K-20 Words :				
K-21 Words :				
K-22 Words :				
K-23 Words :				
K-24 Words :				
K-25 Words :				
Off-List:	??	26 (4.76)	30 (1.98)	98.55
Total (unrounded)	431+?	546 (100)	1515 (100)	≈100.00

RELATED RATIOS & INDICES

Pertaining to whole text

Words in text (tokens):	1515
Different words (types):	546
Type-token ratio:	0.36
Tokens per type:	2.77

Pertaining to onlist only

Tokens:	1485
Types:	520
Families:	431
Tokens per Family :	3.45
Types per Family :	1.21

Current profile (token %)	
K-1 (70.50)	70.50
K-2 (13.07)	83.57
K-3 (7.39)	90.96

K-4 (1.72)	92.68
K-5 (1.19)	93.87
K-6 (0.33)	94.20
K-7 (0.33)	94.53
K-8 (0.26)	94.79
K-9 (0.26)	95.05
K-10 (1.12)	96.17
K-11 (0.20)	96.37
K-14 (0.07)	96.44
K-18 (0.13)	96.57
OFF (1.98)	≈100%

Manual adjustment of the off-list category

Current profile (token %)	
K-1 (72.01)	72.01
K-2 (13.20)	85.21
K-3 (7.46)	92.67
K-4 (1.72)	94.39
K-5 (1.25)	95.64
K-6 (0.40)	96.04
K-7 (0.33)	96.73
K-8 (0.26)	96.99
K-9 (0.26)	97.25
K-10 (1.12)	98.37
K-11 (0.20)	98.57
K-14 (0.07)	98.64
K-18 (0.13)	98.77
OFF (.13)	≈100%

A. Types list

BNC-COCA-1,000 types: [fams 198 : types 241 : tokens 994]

a_[36] about_[3] accepted_[2] according_[1] action_[2] actions_[1]
 additional_[1] address_[1] against_[5] age_[2] all_[4] along_[3] among_[1]
 amongst_[1] an_[8] and_[60] are_[3] areas_[1] as_[16] at_[7] back_[1]
 based_[3] be_[2] become_[2] been_[5] before_[1] began_[2] beginning_[1]
 beyond_[1] blackfoot_[1] bloody_[1] boards_[1] book_[1] books_[1] born_[1]
 bring_[1] brought_[1] but_[1] by_[7] call_[1] called_[1] calling_[2] calls_[1]
 carried_[2] central_[1] change_[9] changed_[1] changes_[1] clear_[2]

close_[1] concern_[1] concerning_[1] considering_[1] continued_[2]
 continues_[1] continuing_[1] country_[3] court_[2] crimes_[1] dances_[1]
 dating_[1] deal_[1] deeply_[1] did_[1] different_[1] do_[1] dropped_[1]
 during_[1] early_[2] educate_[2] education_[1] end_[4] ending_[1]
 especially_[1] ever_[1] experience_[2] fact_[1] fifty_[2] first_[1] following_[1]
 football_[11] for_[19] forced_[1] four_[1] from_[7] gave_[1] general_[1]
 government_[2] governments_[1] great_[1] ground_[1] grown_[2] half_[1]
 hands_[1] has_[12] hate_[1] hateful_[1] have_[8] he_[3] head_[1] hear_[1]
 held_[1] high_[3] his_[1] historian_[1] historical_[1] holds_[1] honorably_[1]
 honoring_[2] however_[1] human_[1] hundreds_[1] hunting_[1] in_[39] is_[6]
 issue_[2] it_[2] its_[4] join_[1] just_[1] largest_[1] last_[2] later_[1] leader_[1]
 leading_[1] level_[1] levels_[2] life_[1] local_[1] long_[3] made_[1] make_[2]
 many_[1] mark_[1] marketing_[2] marks_[1] may_[1] members_[1]
 membership_[1] million_[1] misleading_[1] moment_[1] more_[1] most_[2]
 moved_[1] movement_[2] movies_[1] much_[2] name_[14] names_[1]
 nation_[4] national_[6] nations_[2] nearly_[3] need_[1] new_[1] newer_[1]
 not_[7] number_[30] numbers_[3] of_[63] office_[2] oklahoma_[3] oldest_[1]
 on_[10] once_[1] only_[1] or_[3] other_[1] others_[1] our_[1] out_[1]
 over_[2] owner_[2] owners_[1] page_[1] parts_[1] passed_[1] past_[1]
 people_[11] peoples_[9] personhood_[1] photo_[3] place_[1] players_[2]
 position_[2] post_[2] press_[1] problem_[1] protect_[1] protections_[1]
 public_[4] race_[3] races_[1] rates_[1] rather_[2] reached_[1] real_[2]
 recent_[1] reported_[1] reportedly_[1] reports_[1] rights_[6] ruled_[1]
 rulings_[1] s_[1] school_[2] schools_[1] science_[1] sea_[1] self_[1]
 serious_[1] served_[1] service_[1] seventy_[1] since_[2] so_[1]
 someone_[1] sports_[15] standing_[2] started_[1] still_[1] stronger_[1]
 support_[5] take_[1] taken_[1] taxpayer_[1] team_[18] teams_[2] term_[3]
 terms_[1] than_[4] that_[8] the_[101] their_[4] there_[2] these_[5] this_[3]
 those_[1] through_[1] time_[3] to_[50] today_[3] together_[1] took_[1]
 truly_[1] ugly_[2] under_[2] until_[2] up_[2] us_[1] use_[3] used_[1] very_[2]
 video_[2] voice_[1] want_[1] was_[9] we_[1] well_[4] were_[4] what_[2]
 when_[4] while_[1] who_[1] wide_[2] with_[5] word_[3] working_[1]
 would_[1] year_[1] years_[7] your_[1]

BNC-COCA-2,000 types: [fams 88 : types 96 : tokens 199]

activity_[1] alarmingly_[1] announcements_[1] appeal_[1] appointed_[1]
 association_[1] attention_[1] brand_[2] brands_[1] braves_[1] brief_[1]
 century_[1] challenge_[3] chapter_[1] chief_[1] communities_[1]
 community_[1] contacted_[1] contribute_[1] cultural_[2] culture_[8]
 cultures_[1] current_[1] decision_[1] department_[1] directly_[1]
 disregard_[1] dollar_[1] editorial_[1] effort_[2] efforts_[1] entire_[1]
 equality_[1] established_[1] expense_[1] filed_[1] financed_[1] harm_[2]
 harmful_[11] illustrate_[1] image_[1] images_[2] immediate_[1] including_[3]
 indicate_[1] individuals_[1] interview_[1] justice_[2] lack_[1] language_[1]

league_2] legal_1] likely_1] lower_1] maintain_1] mounted_1]
native_40] non_2] opposite_1] opposition_3] organization_3]
organizations_4] original_2] percent_2] policies_2] popular_5]
population_2] positive_1] practices_1] president_6] pro_2]
professional_2] progress_2] property_1] proud_2] recognized_1]
refused_1] regional_1] register_1] registered_1] removal_1] replace_1]
represent_1] representative_1] representatives_1] research_1]
reservations_1] result_1] rooted_3] selected_1] social_4] society_2]
specifically_1] spread_1] states_1] storm_1] success_1] successful_1]
therefore_1] traditional_1] version_1]

BNC-COCA-3,000 types: [fams 61 : types 68 : tokens 114]

adopted_1] advocate_2] advocates_1] allies_1] analysis_1] ban_1]
banning_1] campaign_6] cancellation_1] ceremonies_1] chairman_1]
civil_5] civilization_1] component_1] comprehensive_1] condemning_1]
confining_1] congress_3] consequences_2] decades_1] defined_1]
derived_1] documented_1] dominant_1] entitled_1] era_4] factors_1]
federal_1] federally_1] former_1] foundation_1] incorporated_2]
institute_1] integrate_1] invisibility_1] launch_1] launched_2] media_6]
myth_1] negative_3] origin_1] origins_1] personalities_1] poses_1]
promoted_1] psychological_1] racial_4] racism_3] racist_2]
regulations_1] reputation_1] request_1] requesting_1] resolution_2]
resolutions_1] retain_1] review_1] significant_2] submit_1]
substantial_1] supreme_1] swell_1] symbols_1] trend_1] tribal_4]
tribe_2] tribes_2] updates_1] urged_1] via_1] violence_1] youth_5]

BNC-COCA-4,000 types: [fams 17 : types 17 : tokens 26]

amateur_1] behalf_1] contrary_1] copyright_1] dictionary_1] equity_1]
franchise_1] franchises_1] helmets_1] legacy_2] opted_1] outlets_1]
patent_2] petition_1] portray_1] savage_1] stereotypes_7] terminate_1]

BNC-COCA-5,000 types: [fams 12 : types 12 : tokens 24]

advocacy_1] assimilate_1] diffuse_1] erase_1] intolerance_1] logo_6]
logos_1] marshall_6] quest_1] segregationist_2] spear_1] stubbornly_1]
uphold_1]

BNC-COCA-6,000 types: [fams 5 : types 5 : tokens 5]

confiscating_1] outlawing_1] perpetuate_1] reverted_1] scalps_1]

BNC-COCA-7,000 types: [fams 5 : types 5 : tokens 5]

bounty_1] caricatures_1] encyclopedia_1] multi_1] viral_1]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 4]

bigotry_1] slur_3]

BNC-COCA-9,000 types: [fams 4 : types 4 : tokens 6]

collegiate_2] denigrating_1] disparaging_1] twitter_2]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 17]

mascot_6] mascots_11]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 3]

derogatory_3]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams 1 : types 1 : tokens 1]

amicus_1]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams 2 : types 2 : tokens 2]

asocial_1] superbowl_1]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 5]

aka_[1] hashtag_[1] heightened_[1] waived_[1] interchangeably_[1]

B. Families list

BNC-COCA-1,000 Families: [fams 196 : types 239 : tokens 982]

a_[44] about_[3] accept_[2] act_[2] add_[1] address_[1] against_[5] age_[2] all_[4] along_[3] among_[2] and_[60] area_[1] as_[15] at_[7] back_[1] base_[3] be_[30] become_[2] before_[1] begin_[3] beyond_[1] blood_[1] board_[1] book_[2] born_[1] bring_[2] but_[1] by_[6] call_[5] carry_[2] centre_[1] change_[11] clear_[2] close_[1] concern_[2] consider_[1] continue_[4] country_[3] court_[2] crime_[1] dance_[1] deal_[1] deep_[1] different_[1] do_[2] drop_[1] during_[1] early_[2] educate_[3] end_[5] end_of_list_[1] especially_[1] ever_[1] experience_[2] fact_[1] first_[1] five_[2] follow_[1] football_[10] for_[19] force_[1] four_[1] from_[7] general_[1] give_[1] govern_[2] great_[1] ground_[1] grow_[2] half_[1] hand_[1] hate_[2] have_[20] he_[4] head_[1] hear_[1] high_[3] history_[2] hold_[2] honour_[3] however_[1] human_[1] hundred_[1] hunt_[1] in_[38] issue_[2] it_[6] join_[1] just_[1] large_[1] last_[2] late_[1] lead_[3] level_[3] life_[1] local_[1] long_[3] make_[3] many_[1] mark_[2] market_[2] may_[1] member_[2] million_[1] moment_[1] more_[1] most_[2] move_[3] movie_[1] much_[2] name_[15] nation_[12] near_[3] need_[1] new_[2] not_[7] number_[33] of_[62] office_[2] old_[1] on_[10] once_[1] only_[1] or_[3] other_[2] out_[1] over_[2] owned_[3] page_[1] part_[1] pass_[1] past_[1] people_[20] person_[1] photograph_[3] place_[1] play_[2] position_[2] post_[2] press_[1] problem_[1] protect_[2] public_[4] race_[4] rate_[1] rather_[2] reach_[1] real_[2] recent_[1] report_[3] rights_[6] rule_[2] school_[3] science_[1] sea_[1] self_[1] serious_[1] serve_[1] service_[1] seven_[1] since_[2] so_[1] some_[1] sport_[15] stand_[2] start_[1] still_[1] strong_[1] support_[5] take_[3] tax_[1] team_[20] term_[4] than_[4] that_[9] the_[101] there_[2] they_[4] this_[8] through_[1] time_[3] to_[50] today_[3] together_[1] true_[1] ugly_[2] under_[2] until_[2] up_[2] use_[4] very_[2] video_[2] voice_[1] want_[1] we_[3] well_[4] what_[2] when_[4] while_[1] who_[1] wide_[2] with_[5] word_[3] work_[1] would_[1] year_[8] you_[1]

BNC-COCA-2,000 Families: [fams 88 : types 96 : tokens 199]

active_[1] alarm_[1] announce_[1] appeal_[1] appoint_[1] associate_[1] attention_[1] brand_[3] brave_[1] brief_[1] century_[1] challenge_[3] chapter_[1] chief_[1] community_[2] contact_[1] contribute_[1] culture_[11] current_[1] decision_[1] department_[1] direct_[1] dollar_[1] edit_[1] effort_[3] entire_[1] equal_[1] establish_[1] expense_[1] file_[1] finance_[1] harm_[13] illustrate_[1] image_[3] immediate_[1] include_[3] indicate_[1]

individual_[1] interview_[1] justice_[2] lack_[1] language_[1] league_[2] legal_[1] likely_[1] lower_[1] maintain_[1] mount_[1] native_[40] non_[2] oppose_[3] opposite_[1] organize_[7] original_[2] percent_[2] policy_[2] popular_[5] population_[2] positive_[1] practise_[1] president_[6] profession_[4] progress_[2] property_[1] proud_[2] recognize_[1] refuse_[1] regard_[1] region_[1] register_[2] remove_[1] replace_[1] represent_[3] research_[1] reserve_[1] result_[1] root_[3] select_[1] social_[4] society_[2] specific_[1] spread_[1] states_[1] storm_[1] success_[2] therefore_[1] tradition_[1] version_[1]

BNC-COCA-3,000 Families: [fams 61 : types 68 : tokens 114]

adopt_[1] advocate_[3] ally_[1] analyse_[1] ban_[2] campaign_[6] cancel_[1] ceremony_[1] chairman_[1] civil_[5] civilise_[1] component_[1] comprehensive_[1] condemn_[1] confine_[1] congress_[3] consequence_[2] decade_[1] define_[1] derive_[1] document_[1] dominant_[1] entitle_[1] era_[4] factor_[1] federal_[2] former_[1] foundation_[1] incorporate_[2] institute_[1] integrate_[1] launch_[3] media_[6] myth_[1] negative_[3] origin_[2] personality_[1] pose_[1] promote_[1] psychology_[1] racial_[9] regulate_[1] reputation_[1] request_[2] resolution_[3] retain_[1] review_[1] significant_[2] submit_[1] substantial_[1] supreme_[1] swell_[1] symbol_[1] trend_[1] tribe_[8] update_[1] urge_[1] via_[1] violence_[1] visible_[1] youth_[5]

BNC-COCA-4,000 Families: [fams 17 : types 17 : tokens 26]

amateur_[1] behalf_[1] contrary_[1] copyright_[1] dictionary_[1] equity_[1] franchise_[2] helmet_[1] legacy_[2] opt_[1] outlet_[1] patent_[2] petition_[1] portray_[1] savage_[1] stereotype_[7] terminate_[1]

BNC-COCA-5,000 Families: [fams 12 : types 12 : tokens 24]

advocacy_[1] assimilate_[1] diffuse_[1] erase_[1] logo_[7] marshal_[6] quest_[1] segregate_[2] spear_[1] stubborn_[1] tolerant_[1] uphold_[1]

BNC-COCA-6,000 Families: [fams 5 : types 5 : tokens 5]

confiscate_[1] outlaw_[1] perpetuate_[1] revert_[1] scalp_[1]

BNC-COCA-7,000 Families: [fams 5 : types 5 : tokens 5]

bounty_[1] caricature_[1] encyclopedia_[1] multi_[1] viral_[1]

BNC-COCA-8,000 Families: [fams 2 : types 2 : tokens 4]

bigot_[1] slur_[3]

BNC-COCA-9,000 Families: [fams 4 : types 4 : tokens 6]

collegiate_[2] denigrate_[1] disparage_[1] twitter_[2]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 17]

mascot_[17]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 3]

derogatory_[3]

BNC-COCA-12,000 Families: [fams : types : tokens]

BNC-COCA-13,000 Families: [fams : types : tokens]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

amicus_[1]

BNC-COCA-15,000 Families: [fams : types : tokens]

BNC-COCA-16,000 Families: [fams : types : tokens]

BNC-COCA-17,000 Families: [fams : types : tokens]

BNC-COCA-18,000 Families: [fams 2 : types 2 : tokens 2]

asocial_[1] superbowl_[1]

BNC-COCA-19,000 Families: [fams : types : tokens]

BNC-COCA-20,000 Families: [fams : types : tokens]

BNC-COCA-21,000 Families: [fams : types : tokens]

BNC-COCA-22,000 Families: [fams : types : tokens]

BNC-COCA-23,000 Families: [fams : types : tokens]

BNC-COCA-24,000 Families: [fams : types : tokens]

BNC-COCA-25,000 Families: [fams : types : tokens]

OFFLIST: [?: types 5 : tokens 5]

aka_[1] hashtag_[1] hightened_[1] waived_[1] interchangeably_[1]

7.3 Text analyses – glossary items

7.3.1. Access to English

7.3.1.1. Divided by a Common Language

Words: distinct, eventually, eavesdrop, host, particularly, spell, feature, witty, proposal, rivalry

Collocations:

succeed in (11 tokens in total)

Vocabprofiler

AWL [3:3:3] distinct eventually feature

Sublist 2
distinct feature

Sublist 8
eventually

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]
particularly_[1]

BNC-COCA-2,000 types: [fams 4 : types 4 : tokens 4]

eventually_[1] feature_[1] proposal_[1] spell_[1]

BNC-COCA-3,000 types: [fams 4 : types 4 : tokens 4]

distinct_[1] host_[1] rivalry_[1] succeed_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

witty_[1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

[eavesdrop_\[1\]](#)

7.3.1.2 *A Global Language*

Words:

edge, defy, invent, accession, reign, encourage, foothold, domestic, far-flung, sparse, disunited, defeat, remain, bi-lingual, reveal, barren, inhospitable, aborigine, convict, warrior, sheer, option, tiny, manpower, scientist, aviation, knitting, negotiate, ignore, advert, prawn, predict, impact

Collocations:

vast majority, leader of the pack, due to, rely on, in charge (38 tokens in total)

Lextutor – Vocabprofiler

BNC-COCA-1,000 types: [fams 12 : types 12 : tokens 12]

BNC-COCA-1,000 types: [fams 6 : types 6 : tokens 6]

[advert_\[1\]](#) [charge_\[1\]](#) [edge_\[1\]](#) [leader_\[1\]](#) [pack_\[1\]](#) [scientist_\[1\]](#)

BNC-COCA-2,000 types: [fams 8 : types 8 : tokens 8]

[due_\[1\]](#) [encourage_\[1\]](#) [ignore_\[1\]](#) [knitting_\[1\]](#) [option_\[1\]](#) [rely_\[1\]](#) [remain_\[1\]](#)
[tiny_\[1\]](#)

BNC-COCA-3,000 types: [fams 9 : types 9 : tokens 9]

[convict_\[1\]](#) [defeat_\[1\]](#) [domestic_\[1\]](#) [impact_\[1\]](#) [invent_\[1\]](#) [vast majority_\[1\]](#)
[negotiate_\[1\]](#) [predict_\[1\]](#) [reveal_\[1\]](#)

BNC-COCA-4,000 types: [fams 4 : types 4 : tokens 4]

[defy_\[1\]](#) [reign_\[1\]](#) [sheer_\[1\]](#) [warrior_\[1\]](#)

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

[aviation_\[1\]](#)

BNC-COCA-6,000 types: [fams 2 : types 2 : tokens 2]

[barren_\[1\]](#) [sparse_\[1\]](#)

BNC-COCA-7,000 types: [fams 2 : types 2 : tokens 2]

[aborigine_\[1\]](#) [bilingual_\[1\]](#)

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

[accession_\[1\]](#) [inhospitable_\[1\]](#)

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

[prawn_\[1\]](#)

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 4 : tokens 4]

[disunited_\[1\]](#) [farflung_\[1\]](#) [foothold_\[1\]](#) [manpower_\[1\]](#)

7.3.1.3 Native Americans-Original Inhabitants

Words:

migrate, diverse, hide, pre-Colombian, thrive, adobe, remain, dwelling, cliff, disastrous, immunity, smallpox, decline, disease, convert, undermine, ban, hostile, benefit, access, defeat, backwoods, disrupted, relocate, rare, notable, disobey, foolhardy, estimate, record, casualty, inequality, heritage,

Collocations:

at the expense of (34 tokens in total)

Lextutor – Vocabprofiler

AWL [9:9:9] access benefit convert decline diverse estimate migrate relocate require

Sublist 1

benefit estimate require

Sublist 3

relocate

Sublist 4

access

Sublist 5

decline

Sublist 6

diverse migrate

Sublist 7

convert

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

hide_[1] notable_[1] record_[1]

BNC-COCA-2,000 types: [fams 10 : types 10 : tokens 10]

access_[1] benefit_[1] cliff_[1] disease_[1] expense_[1] inequality_[1] rare_[1] relocate_[1] remain_[1] require_[1]

BNC-COCA-3,000 types: [fams 13 : types 13 : tokens 13]

aggressive_[1] ban_[1] convert_[1] decline_[1] defeat_[1] disastrous_[1] disrupted_[1] diverse_[1] estimate_[1] heritage_[1] hostile_[1] immunity_[1] migrate_[1]

BNC-COCA-4,000 types: [fams 4 : types 4 : tokens 4]

casualty_[1] dwelling_[1] thrive_[1] undermine_[1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

pre_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

adobe_[1]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

smallpox_[1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

disobey_[1]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

foolhardy_[1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

backwood_[1] colombian_[1]

7.3.1.4 Aboriginal Australians

Glossing – Access to English: Aboriginal Australians

Words:

uncharted, musket, ashore, spear, telling, explorer, ancestor, intruder, ignore, convict, teeming, far-flung, uninhabitable, track, stalk, lizard, emerge, temperate, territorial, distress, mystified, apparent, creation, accessible, pest, dingo, nomadic, stay, providence, resistance, ambush, coordination, measles, chickenpox, smallpox, wildfire, game, humiliation, skeleton, exhibit, adapt, remote, segregation, issue, admit, contest, dismal, sovereignty, claim, penal, concept, deprive, domesticated, account, prey, lag

Collocations:

claim for, penal colony, no concept of, deprive of, domesticated animal, account for, wipe out, fall prey to, lag behind,

Lextutor – Vocabprofiler

AWL [12:12:12] accessible adapt apparent co-ordination concept creation domesticated emerge exhibit ignore issue uncharted

Sublist 1

concept creation issue

Sublist 3

coordination

Sublist 4

accessible apparent domesticated emerge

Sublist 6

ignore

Sublist 7

adapt

Sublist 8

exhibit uncharted

BNC-COCA-1,000 types: [fams 8 : types 8 : tokens 8]

admit_[1] apparent_[1] far_[1] game_[1] issue_[1] stay_[1] telling_[1] track_[1]

BNC-COCA-2,000 types: [fams 7 : types 7 : tokens 7]

accessible_[1] account_[1] adapt_[1] claim_[1] creation_[1] ignore_[1] resistance_[1]

BNC-COCA-3,000 types: [fams 12 : types 12 : tokens 12]

concept_[1] contest_[1] convict_[1] coordination_[1] emerge_[1] exhibit_[1] explorer_[1] remote_[1] sovereignty_[1] territorial_[1] uncharted_[1] uninhabitable_[1]

BNC-COCA-4,000 types: [fams 6 : types 6 : tokens 6]

ancestor_[1] deprive_[1] distress_[1] flung_[1] humiliation_[1] prey_[1]

BNC-COCA-5,000 types: [fams 7 : types 7 : tokens 7]

intruder_[1] lag_[1] pest_[1] segregation_[1] skeleton_[1] spear_[1] stalk_[1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

ambush_[1] ashore_[1] dismal_[1] lizard_[1]

BNC-COCA-7,000 types: [fams 3 : types 3 : tokens 3]

nomadic_[1] penal_[1] temperate_[1]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

domesticates_[1] mystified_[1]

BNC-COCA-9,000 types: [fams 3 : types 3 : tokens 3]

musket_[1] smallpox_[1] teeming_[1]

BNC-COCA-10,000 types: [fams 2 : types 2 : tokens 2]

measles_[1] providence_[1]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens 1]

dingo_[1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 2 : tokens 2]

chickenpox_[1] wildfire_[1]

7.3.1.5 Stolen Children

Words: assimilate, uprooted, outcry, issue, reconciliation, orphanage

Collocations: mainstream society

Lextutor – Vocabprofiler

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]

issue_[1]

BNC-COCA-2,000 types: [fams 1 : types 1 : tokens 1]

society_[1]

BNC-COCA-3,000 types: [fams : types : tokens]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

reconciliation_[1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

assimilate_[1] orphanage_[1]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

outcry_[1] uprooted_[1]

OFFLIST: [?: types 1 : tokens 1]

mainstream_[1]

7.3.2 Stunt

7.3.2.1 British vs. American English

Words: divided, common, conscious, consistent, evident, gregarious, pronunciation, analyse, syllable, punctuation, profanity

Collocations: -

Lextutor – Vocabprofiler

AWL [3:3:3] analyse consistent evident

Sublist 1
analyse consistent evident

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 0 : types 0 : tokens 0]

BNC-COCA-2,000 types: [fams 4 : types 4 : tokens 4]

common_[1] conscious_[1] divided_[1] pronunciation_[1]

BNC-COCA-3,000 types: [fams 3 : types 3 : tokens 3]

analyse_[1] consistent_[1] evident_[1]

BNC-COCA-4,000 types: [fams : types : tokens]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 2 : types 2 : tokens 2]

punctuation_[1] syllable_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

profanity_[1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

gregarious_[1]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.2.2 *English as a World Language*

Words: interaction, means, reign, experience, impact, coupled, turmoil, exposed, participate,

Collocations:

coupled with, exposed to

Note: The token *impact* is represented in the glossary, but is not found in the text. It has therefore been removed from the analysis.

Lextutor – Vocabprofiler

AWL [4:4:4] coupled exposed interaction participate

Sublist 2

impact participate

Sublist 3

interaction

Sublist 5

exposed

Sublist 7

coupled

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

[coupled](#)_[1] [experience](#)_[1] [means](#)_[1]

BNC-COCA-2,000 types: [fams 1 : types 1 : tokens 1]

[exposed](#)_[1]

BNC-COCA-3,000 types: [fams 2 : types 2 : tokens 2]

[interaction](#)_[1] [participate](#)_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

[reign](#)_[1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

[turmoil](#)_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.2.3 Native Americans

Words: Indigenous, conquer, tumultuous, strained, policy, removal, access, perish, cultivate, descent

Collocations: -

Lextutor – Vocabprofiler

AWL [3:3:3] access policy removal

Sublist 1

policy

Sublist 3

removal

Sublist 4

access

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 0 : types 0 : tokens 0]

BNC-COCA-2,000 types: [fams 3 : types 3 : tokens 3]

[access_\[1\]](#) [policy_\[1\]](#) [removal_\[1\]](#)

BNC-COCA-3,000 types: [fams 1 : types 1 : tokens 1]

[strained_\[1\]](#)

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

[cultivate_\[1\]](#) [indigenous_\[1\]](#)

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

[conquer_\[1\]](#) [descent_\[1\]](#)

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

[perish_\[1\]](#)

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

[tumultuous_\[1\]](#)

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.2.4 Australia: The Birth of a Nation

Words: Inhabited, consideration, explorers, claim, convict, cultivate, fortune, encourage, surplus, referendum

Collocations: fortune hunters

Lexturor – Vocabprofiler

AWL – 0

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]

[consideration_\[1\]](#)

BNC-COCA-2,000 types: [fams 3 : types 3 : tokens 3]

[claim_\[1\]](#) [encourage_\[1\]](#) [fortune_\[1\]](#)

BNC-COCA-3,000 types: [fams 3 : types 3 : tokens 3]

[convict_\[1\]](#) [explorers_\[1\]](#) [inhabited_\[1\]](#)

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

[cultivate_\[1\]](#) [surplus_\[1\]](#)

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

[referendum_\[1\]](#)

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.2.5 Stolen Generation

Words:

Circumstances, smallpox, fertile, inferior, perceive, orphanages, neglect, abuse, estimated, apology, predecessor

Collocations:

Proper nouns:

Torres Strait Islands

Lexturor – Vocabprofiler

AWL [3:3:3] circumstances estimated perceive

Sublist 1
estimated

Sublist 2
perceive

Sublist 3
circumstances

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 0 : types 0 : tokens 0]

BNC-COCA-2,000 types: [fams 1 : types 1 : tokens 1]

[circumstances_\[1\]](#)

BNC-COCA-3,000 types: [fams 6 : types 6 : tokens 6]

[abuse_\[1\]](#) [apology_\[1\]](#) [estimated_\[1\]](#) [fertile_\[1\]](#) [neglect_\[1\]](#) [perceive_\[1\]](#)

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

[predecessor_\[1\]](#)

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

[inferior_\[1\]](#) [orphanages_\[1\]](#)

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 1 : types 1 : tokens 1]

[smallpox_\[1\]](#)

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.3 Targets

7.3.3.1 *The Flavours of English*

Words: regard, drawl, means, distinct, features, rhotic, ties, convict, trace, respects, obvious, descent, aspiration, trilled, posh, expansion, emergence, boondoggle, kilter, diphthong, nasalized, broad, abbreviations

Collocations: with regard to, by no means, in many respects, native tongue, tag questions, out of kilter

Proper nouns:

cockney, received pronunciation, pidgin, creole

Lextutor – Vocabprofiler

AWL [6:6:6] distinct emergence expansion features obvious trace

Sublist 2

distinct features

Sublist 4

emergence obvious

Sublist 5

expansion

Sublist 6

trace

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

[means_\[1\]](#) [obvious_\[1\]](#) [ties_\[1\]](#)

BNC-COCA-2,000 types: [fams 5 : types 5 : tokens 5]

[broad_\[1\]](#) [features_\[1\]](#) [regard_\[1\]](#) [respects_\[1\]](#) [trace_\[1\]](#)

BNC-COCA-3,000 types: [fams 4 : types 4 : tokens 4]

[convict_\[1\]](#) [distinct_\[1\]](#) [emergence_\[1\]](#) [expansion_\[1\]](#)

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

aspiration_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

descent_[1]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams 3 : types 3 : tokens 3]

abbreviations_[1] drawl_[1] nasalized_[1]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

posh_[1]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

trilled_[1]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 2 : types 2 : tokens 2]

diphthong_[1] kilter_[1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams 1 : types 1 : tokens 1]

boondoggle_[1]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 1 : tokens 1]

rhotic_[1]

7.3.3.2 The Power of English Part 1

Words: opportunity, twofold, influence, empower, prospects, increase, estimate, presumably, furthermore, guesstimate, outnumber, mutually, intelligible, account, uncertainty, primary, yet, sovereign, non-sovereign, entity, nevertheless, justify, enable, extensively, commerce, further, contribute, promote

Collocations:

native tongue, take into account, contribute to,

Note: The following words are in the glossary, but not used in the text: *fluency, prevailing, summit, unintelligible* and *emergency services*. They have therefore been removed from the analysis.

Lextutor – Vocabprofiler

AWL [12:12:12] contribute enable entity estimate furthermore justify mutually nevertheless presumably primary promote prospects

Sublist 1
estimate

Sublist 2
primary

Sublist 3
contribute justify

Sublist 4
promote

Sublist 5
enable entity

Sublist 6

furthermore nevertheless presumably

Sublist 8

prospects

Sublist 9

mutually

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

further_[1] uncertainty_[1] yet_[1]

BNC-COCA-2,000 types: [fams 7 : types 7 : tokens 7]

account_[1] commerce_[1] contribute_[1] increase_[1] influence_[1]
opportunity_[1] presumably_[1]

BNC-COCA-3,000 types: [fams 11 : types 11 : tokens 12]

enable_[1] estimate_[1] extensively_[1] furthermore_[1] justify_[1]
mutually_[1] nevertheless_[1] primary_[1] promote_[1] prospects_[1]
sovereign_[2]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

entity_[1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

empower_[1] intelligible_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

twofold_[1]

BNC-COCA-7,000 types: [fams 1 : types 1 : tokens 1]

outnumber_[1]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams 1 : types 1 : tokens

guesstimate_[1]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.3.3.3 The Power of English Part 2

Words: Unravel, podium, exert, expand, settlement, trade, gain, attempt, far-reaching, reverberate, headway, claim, post, primarily, judicial, lucrative, spices, fierce, populous, onset, outpost, cornerstone, fleet, seep, convict, nook, voyage, hazardous, launch, outnumber, adventurous, interior, scramble, maintain, asset, protectorate, booming, expanding, boost, coalition, treaty, peak, immense, unrivalled, root, workshop, dismantle, visible, exposed, supremacy, suffice, setback

Collocations: exert power, gain ground, attempt at, make headway, trading post, merchant fleet, seep into, nooks and crannies, sea voyage, pick up, the interior, peace treaty, at its peak, suffice to say,

Proper nouns:

GDP

Used in the glossary, but not found in the text and therefore removed from the analysis: *take root, decolonization, dominion, unleash, mandate, and enforce.*

Lextutor – Vocabprofiler

AWL [6:7:7] enforce expand expanding exposed maintain primarily visible

Sublist 2

maintain primarily

Sublist 5

enforce expand expanding exposed

Sublist 7

visible

BNC-COCA-1,000 types: [fams 4 : types 4 : tokens 4]

far_[1] post_[1] reaching_[1] settlement_[1]

BNC-COCA-2,000 types: [fams 9 : types 9 : tokens 9]

adventurous_[1] attempt_[1] booming_[1] claim_[1] exposed_[1] gain_[1]
maintain_[1] root_[1] trade_[1]

BNC-COCA-3,000 types: [fams 15 : types 16 : tokens 16]

asset_[1] boost_[1] coalition_[1] convict_[1] enforce_[1] expand_[1]
expanding_[1] fierce_[1] hazardous_[1] interior_[1] launch_[1] peak_[1]
primarily_[1] treaty_[1] unrivalled_[1] visible_[1]

BNC-COCA-4,000 types: [fams 7 : types 7 : tokens 7]

exert_[1] fleet_[1] immense_[1] judicial_[1] mandate_[1] scramble_[1]
spices_[1]

BNC-COCA-5,000 types: [fams 4 : types 4 : tokens 4]

dismantle_[1] onset_[1] unleash_[1] voyage_[1]

BNC-COCA-6,000 types: [fams 4 : types 4 : tokens 4]

lucrative_[1] seep_[1] suffice_[1] unravel_[1]

BNC-COCA-7,000 types: [fams 4 : types 4 : tokens 4]

dominion_[1] outnumber_[1] reverberate_[1] supremacy_[1]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

outpost_[1] podium_[1]

BNC-COCA-9,000 types: [fams 2 : types 2 : tokens 2]

headway_[1] nook_[1]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

populous_[1]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

protectorate_[1]

BNC-COCA-12,000 types: [fams 1 : types 1 : tokens 1]

decolonization_[1]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 3 : tokens 3]

[cornerstone_\[1\]](#) [setback_\[1\]](#) [workshop_\[1\]](#)

7.3.3.4 Native Americans: We Are Still Here

Words: Ancestor, prior, roam, seasonal, significant, descendant, indigenous, smallpox, legacy, treaty, confine, assimilate, constraint, persevere, advocate, vibrant, feathered, headdress, beaded, buckskin, conveniences, plumbing, disadvantage, disproportionately, sovereignty, sovereign, body, contemporary, sacred, artifact, profit, persistent, promote, appearance, vanish

Collocations:

Prior to, seasonal pattern, pay significant attention to, indigenous people, advocate for

The token *sedentary* is in the glossary, but not in the text and has been removed from the analysis.

Lextutor – Vocabprofiler

AWL [9:9:9] [advocate](#) [confine](#) [constraint](#) [contemporary](#) [disproportionately](#) [persistent](#) [prior](#) [promote](#) [significant](#)

Sublist 1
[significant](#)

Sublist 3
[constraint](#) [disproportionately](#)

Sublist 4
[prior](#) [promote](#)

Sublist 7
[advocate](#)

Sublist 8
[contemporary](#)

Sublist 9
[confine](#)

Sublist 10
[persistent](#)

Types List [\[1\]](#)

type_[number of tokens]

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]

[appearance_\[1\]](#) [body_\[1\]](#)

BNC-COCA-2,000 types: [fams 2 : types 2 : tokens 2]

[feathered_\[1\]](#) [seasonal_\[1\]](#)

BNC-COCA-3,000 types: [fams 12 : types 12 : tokens 13]

[advocate_\[1\]](#) [confine_\[1\]](#) [constraint_\[1\]](#) [contemporary_\[1\]](#) [descendant_\[1\]](#) [persistent_\[1\]](#) [prior_\[1\]](#) [profit_\[1\]](#) [promote_\[1\]](#) [significant_\[1\]](#) [sovereign_\[1\]](#) [sovereignty_\[1\]](#) [treaty_\[1\]](#)

BNC-COCA-4,000 types: [fams 8 : types 8 : tokens 8]

[ancestor_\[1\]](#) [beaded_\[1\]](#) [conveniences_\[1\]](#) [disadvantage_\[1\]](#) [indigenous_\[1\]](#) [legacy_\[1\]](#) [sacred_\[1\]](#) [vanish_\[1\]](#)

BNC-COCA-5,000 types: [fams 3 : types 3 : tokens 3]

[assimilate_\[1\]](#) [plumbing_\[1\]](#) [roam_\[1\]](#)

BNC-COCA-6,000 types: [fams 2 : types 2 : tokens 2]

[disproportionately_\[1\]](#) [vibrant_\[1\]](#)

BNC-COCA-7,000 types: [fams 2 : types 2 : tokens 2]

[artifact_\[1\]](#) [persevere_\[1\]](#)

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams 2 : types 2 : tokens 2]

sedentary_[1] smallpox_[1]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams 1 : types 1 : tokens 1]

buckskin_[1]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 1 : tokens 1]

headdress_[1]

7.3.3.5 Australia – the Island Continent

Words: Species, hemisphere, ancient, indigenous, undisturbed, descendant, penal, exploited, recognized, influx, clustered, dispersed, vast, idiosyncratic, laconic, revive, maintain, ensure, autonomy, nominal, sentiment, aspiration, compulsory, scale, scope, range, interior, vulnerable, deforestation, overgrazing, cane, feral, sustainable, limited, consequently, loss, drought, supply,

Collocations: desert interior, cane toad, feral cat

Words in the glossary, but not in the text that have been removed from the analysis: *irrigation* and *crop*.

Lextutor – Vocabprofiler

AWL [7:7:7] consequently ensure exploited maintain range scope sustainable

Sublist 2
consequently maintain range

Sublist 3
ensure

Sublist 5
sustainable

Sublist 6
scope

Sublist 8
exploited

BNC-COCA-1,000 types: [fams 0 : types 0 : tokens 0]

BNC-COCA-2,000 types: [fams 9 : types 9 : tokens 9]

limited_[1] loss_[1] maintain_[1] range_[1] recognized_[1] scale_[1]
species_[1] supply_[1] undisturbed_[1]

BNC-COCA-3,000 types: [fams 12 : types 12 : tokens 12]

ancient_[1] clustered_[1] consequently_[1] descendant_[1] ensure_[1]
exploited_[1] interior_[1] revive_[1] scope_[1] sustainable_[1] vast_[1]
vulnerable_[1]

BNC-COCA-4,000 types: [fams 5 : types 5 : tokens 5]

aspiration_[1] autonomy_[1] dispersed_[1] indigenous_[1] sentiment_[1]

BNC-COCA-5,000 types: [fams 4 : types 4 : tokens 4]

cane_[1] drought_[1] hemisphere_[1] nominal_[1]

BNC-COCA-6,000 types: [fams 2 : types 2 : tokens 2]

compulsory_[1] influx_[1]

BNC-COCA-7,000 types: [fams 1 : types 1 : tokens 1]

penal_[1]

BNC-COCA-8,000 types: [fams 2 : types 2 : tokens 2]

deforestation_[1] idiosyncratic_[1]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams 1 : types 1 : tokens 1]

feral_[1]

BNC-COCA-11,000 types: [fams 1 : types 1 : tokens 1]

laconic_[1]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams 1 : types 1 : tokens 1]

overgrazing_[1]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.4 Range analyses

7.4.1 Access to English

7.4.1.1 Global English

Range for Texts - Output

Current analysis title: **Access to English - Topic: Global Language** Language: **English**

INPUT FILES: 3 | FAMS: 785

T_1. (12307 bytes) a global language.txt
T_2. (6746 bytes) divided by a common language.txt
T_3. (4844 bytes) Brisbane times - global english access website.txt
STOPLISTS=1 2 3

000.	Types	Freq	Range	K-BNC	Found in these texts.....
381.	academy	1	1	4	T_3
384.	acknowledge	1	1	4	T_3
387.	administer	1	1	5	T_1
398.	alternative	1	1	2	T_1
406.	area	1	1	1	T_2
413.	aspect	1	1	2	T_1
417.	aware	1	1	1	T_2
427.	benefit	1	1	1	T_3
440.	category	1	1	2	T_1
142.	communicate	4	3	2	T_1 T_2 T_3

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]
 area_[1] aware_[1]

BNC-COCA-2,000 types: [fams 12 : types 12 : tokens 12]

benefit_[1] drama_[1] feature_[1] fund_[1] minor_[1] option_[1] period_[1]
 region_[1] rely_[1] role_[1] site_[1] vary_[1]

BNC-COCA-3,000 types: [fams 29 : types 29 : tokens 29]

academy_[1] acknowledge_[1] alternative_[1] aspect_[1] category_[1]
 consult_[1] distinct_[1] diverse_[1] domestic_[1] element_[1] emphasis_[1]
 error_[1] evolve_[1] export_[1] interact_[1] lecture_[1] link_[1] logic_[1]
 media_[1] participate_[1] perspective_[1] precede_[1] predict_[1]
 proportion_[1] relevant_[1] reveal_[1] source_[1] strategy_[1] structure_[1]

BNC-COCA-4,000 types: [fams 4 : types 4 : tokens 4]

administer_[1] contrary_[1] immigrate_[1] norm_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

convene_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

7.4.1.2 Indigenous Peoples

000.	Fams	Freq	Range	BNCoca	Found in these texts... ..
040.	community	11	3	2	T_1 T_2 T_3
052.	area	9	2	1	T_1 T_2
077.	economy	7	2	2	T_1 T_3
089.	conflict	6	2	3	T_1 T_2
116.	finance	5	1	2	T_3
172.	remove	4	3	2	T_1 T_2 T_4
146.	contact	4	2	2	T_1 T_2
161.	individual	4	2	2	T_1 T_3
162.	issue	4	2	1	T_1 T_4
177.	survive	4	2	2	T_1 T_2
205.	estimate	3	3	3	T_1 T_2 T_3
199.	contribute	3	2	2	T_2 T_3
201.	decline	3	2	3	T_1 T_2

250.	technology	3	2	2	T_1 T_2
190.	assist	3	1	2	T_3
218.	invest	3	1	3	T_3
223.	locate	3	1	2	T_2
385.	source	2	3	3	T_1 T_2 T_3
260.	access	2	2	2	T_1 T_2
266.	approximate	2	2	3	T_1 T_2
278.	civil	2	2	3	T_1 T_2
287.	create	2	2	2	T_1 T_3
288.	culture	2	2	2	T_1 T_2
300.	establish	2	2	2	T_1 T_2
397.	tradition	2	2	2	T_2 T_3
261.	adapt	2	1	2	T_1
282.	convert	2	1	3	T_2
284.	corporate	2	1	3	T_3
306.	feature	2	1	2	T_2
315.	generation	2	1	2	T_4
334.	isolate	2	1	3	T_1
354.	policy	2	1	2	T_1
367.	resource	2	1	3	T_3
391.	style	2	1	2	T_1
594.	grant	1	2	2	T_1 T_4
737.	range	1	2	2	T_1 T_3
418.	administration	1	1	3	T_3
433.	alternative	1	1	3	T_1
442.	apparent	1	1	1	T_1
448.	aspect	1	1	3	T_4
453.	attain	1	1	4	T_3
460.	benefit	1	1	2	T_3
486.	chart	1	1	3	T_1
495.	concentrate	1	1	2	T_2
496.	concept	1	1	3	T_1
501.	coordinate	1	1	3	T_1
508.	debate	1	1	3	T_1
527.	discriminate	1	1	3	T_1
531.	displace	1	1	4	T_2
534.	diverse	1	1	3	T_2
536.	domesticate	1	1	8	T_1
543.	emerge	1	1	3	T_1
547.	enhance	1	1	3	T_3
548.	enormous	1	1	2	T_1
551.	environment	1	1	2	T_2
559.	exclude	1	1	3	T_1
560.	exhibit	1	1	3	T_1
561.	expand	1	1	3	T_2
570.	final	1	1	1	T_2
579.	focus	1	1	3	T_3
585.	fundamental	1	1	3	T_2
590.	goal	1	1	2	T_3
615.	identify	1	1	2	T_1

618.	immigrant	1	1	3		T_4
620.	impact	1	1	3		T_3
624.	innovate	1	1	3		T_3
627.	involve	1	1	1	T_1	
629.	job	1	1	1	T_1	
633.	labour	1	1	2	T_1	
645.	link	1	1	3	T_1	
652.	major	1	1	1		T_3
658.	maximise	1	1	4		T_3
661.	media	1	1	3		T_4
663.	migrate	1	1	3		T_2
664.	military	1	1	2		T_2
682.	obvious	1	1	1	T_1	
683.	occur	1	1	2		T_2
697.	period	1	1	2	T_1	
699.	phenomenon	1	1	3	T_1	
700.	physical	1	1	2	T_1	
709.	potential	1	1	3		T_3
713.	precise	1	1	3	T_1	
719.	priority	1	1	3		T_2
721.	process	1	1	2	T_1	
736.	radical	1	1	3		T_2
742.	react	1	1	2		T_4
751.	region	1	1	2		T_2
801.	stable	1	1	2		T_3
804.	statistic	1	1	3		T_3
805.	status	1	1	3	T_1	
815.	structure	1	1	3		T_2
825.	team	1	1	1		T_2
826.	technique	1	1	3	T_1	
844.	trace	1	1	2		T_2
862.	vary	1	1	2	T_1	

4 | FAMS: 881 |

INPUT CLASSIFIABLE TOKENS:

FILES: 3,751

BNC-COCA-1,000 types: [fams 7 : types 7 : tokens 7]

[apparent](#) [1] [final](#) [1] [involve](#) [1] [job](#) [1] [major](#) [1] [obvious](#) [1] [team](#) [1]

BNC-COCA-2,000 types: [fams 19 : types 19 : tokens 19]

[benefit](#) [1] [concentrate](#) [1] [enormous](#) [1] [environment](#) [1] [goal](#) [1]
[grant](#) [1] [identify](#) [1] [labour](#) [1] [military](#) [1] [occur](#) [1] [period](#) [1] [physical](#) [1]
[process](#) [1] [range](#) [1] [react](#) [1] [region](#) [1] [stable](#) [1] [trace](#) [1] [vary](#) [1]

BNC-COCA-3,000 types: [fams 31 : types 31 : tokens 31]

[administration](#) [1] [alternative](#) [1] [aspect](#) [1] [chart](#) [1] [concept](#) [1]
[coordinate](#) [1] [debate](#) [1] [discriminate](#) [1] [diverse](#) [1] [emerge](#) [1]
[enhance](#) [1] [exclude](#) [1] [exhibit](#) [1] [expand](#) [1] [focus](#) [1] [fundamental](#) [1]
[immigrant](#) [1] [impact](#) [1] [innovate](#) [1] [link](#) [1] [media](#) [1] [migrate](#) [1]
[phenomenon](#) [1] [potential](#) [1] [precise](#) [1] [priority](#) [1] [radical](#) [1] [statistic](#) [1]
[status](#) [1] [structure](#) [1] [technique](#) [1]

BNC-COCA-4,000 types: [fams 3 : types 3 : tokens 3]

[attain](#) [1] [displace](#) [1] [maximise](#) [1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

[domesticate](#) [1]

7.4.2 Stunt

7.4.2.1 Global Language

Range for Texts - Output

Current analysis title: **Stunt - Topic: English as a Universal**

Language: English

INPUT FILES: 3 | FAMS: 571

T_1. (3768 bytes) English as a World Language.txt
T_2. (7006 bytes) british vs american english.txt
T_3. (5643 bytes) An epidemic is threatening.txt

STOPLISTS=3

1. AWL families lists:

a. Unesco – An epidemic

[achieve](#) [1] [aid](#) [1] [circumstance](#) [1] [communicate](#) [2] [community](#) [2]
[complex](#) [1] [concept](#) [2] [constant](#) [1] [constitute](#) [1] [construct](#) [1]
[contact](#) [1] [context](#) [2] [contrary](#) [1] [create](#) [3] [culture](#) [1] [decade](#) [2]

define_[1] despite_[1] estimate_[1] factor_[2] function_[1] gender_[1]
 generation_[3] globe_[2] identify_[1] image_[1] individual_[1] initiate_[4]
 institute_[2] integrate_[1] involve_[1] isolate_[2] maintain_[2] media_[1]
 migrate_[1] minor_[1] policy_[1] predominant_[1] prime_[1] radical_[1]
 region_[4] specific_[2] status_[2] survive_[2] trend_[2] unique_[1] vary_[1]

AWL Fr non-cognate families: [families 2 : tokens 3] involve_[1] trend_[2]

b. British vs. American English

AWL families: [17:19:26]

analyse_[3] area_[1] aspect_[1] category_[1] compute_[1] consist_[2]
 distinct_[1] emphasis_[3] evident_[1] instance_[1] intelligence_[1] quote_[5]
 source_[1] stress_[1] sum_[1] tense_[1] vary_[1]

c. English as a World Language

AWL families: [18:19:22]

communicate_[4] constant_[1] couple_[1] culture_[1] dominate_[1]
 ensure_[1] establish_[1] expose_[2] interact_[1] isolate_[1] job_[1] major_[1]
 minimum_[1] participate_[1] require_[1] role_[1] status_[1] vary_[1]

Output exports to Excel for further manipulation - sort by Frequency,
 Range or Text (default = Freq)

000. Fams Freq Range K-BNC Found in these texts.....

065.	communicate	6	2	2	T_1	T_3
089.	quote	5	1	2		T_2
103.	initiate	4	1	2		T_3
116.	region	4	1	1		T_3
124.	analyse	3	1	2		T_2
129.	continue	3	2	1	T_1	T_3
130.	create	3	1	1		T_3
133.	emphasis		3	1	2	T_2
136.	expose	2	1	3	T_1	

138.	generate	3	1	2		T_3
144.	isolate	3	2	4	T_1	T_3
159.	status	3	2	2	T_1	T_3
195.	community		2	1	1	T_3
196.	concept	2	1	2		T_3
200.	consistent	2	1	2		T_2
202.	context	2	1	2		T_3
206.	culture	2	2	2	T_1	T_3
208.	decade	2	1	4		T_3
218.	factor	2	1	2		T_3
222.	globe	2	1	4		T_3
231.	job	2	1	1	T_1	
240.	maintain	2	1	2		T_3
262.	role	2	1	1	T_1	
273.	specific	2	1	1		T_3
274.	survive	2	1	2		T_3
284.	trend	2	1	4		T_3
297.	achieve	1	1	1		T_3
301.	aid	1	1	2		T_3
305.	area	1	1	1		T_2
307.	aspect	1	1	2		T_2
325.	category	1	1	2		T_2
328.	circumstance		1	1	2	T_3
335.	complex		1	1	2	T_3
337.	compute		1	1	1	T_2
341.	consonant		1	1	10	T_2
342.	constitution		1	1	5	T_3
343.	construct	1	1	4		T_3
344.	contact	1	1	1		T_3
345.	contrary	1	1	5		T_3

347.	couple	1	1	1	T_1	
350.	define	1	1	2		T_3
351.	despite	1	1	2		T_3
354.	distinct	1	1	2		T_2
355.	dominate		1	1	3	T_1
364.	ensure	1	1	2	T_1	
366.	establish	1	1	2	T_1	
367.	estimate	1	1	2		T_3
384.	function	1	1	1		T_3
401.	identify	1	1	1		T_3
402.	image	1	1	2		T_3
407.	individual	1	1	1		T_3
408.	instance	1	1	2		T_2
411.	integrate	1	1	4		T_3
412.	intelligence		1	1	2	T_2
413.	interact	1	1	4	T_1	
416.	involve	1	1	1		T_3
435.	major	1	1	1	T_1	
441.	media	1	1	4		T_3
443.	migrate	1	1	5		T_3
444.	minimum		1	1	2	T_1
445.	minor	1	1	2		T_3
461.	participate	1	1	4	T_1	
462.	particular	1	1	1		T_2
473.	policy	1	1	1		T_3
477.	predominant	1	1	5		T_3
482.	prime	1	1	2		T_3
490.	radical	1	1	6		T_3
501.	require	1	1	1	T_1	
516.	source	1	1	2		T_2

527.	stress	1	1	2		T_2
530.	sum	1	1	2		T_2
536.	tense	1	1	3		T_2
554.	unique	1	1	3		T_3

BNC-COCA-1,000 types: [fams 6 : types 6 : tokens 6]

[area_\[1\]](#) [compute_\[1\]](#) [couple_\[1\]](#) [involve_\[1\]](#) [major_\[1\]](#) [particular_\[1\]](#)

BNC-COCA-2,000 types: [fams 14 : types 14 : tokens 14]

[aid_\[1\]](#) [circumstance_\[1\]](#) [contact_\[1\]](#) [establish_\[1\]](#) [identify_\[1\]](#) [image_\[1\]](#)
[individual_\[1\]](#) [instance_\[1\]](#) [minor_\[1\]](#) [policy_\[1\]](#) [prime_\[1\]](#) [require_\[1\]](#)
[stress_\[1\]](#) [tense_\[1\]](#)

BNC-COCA-3,000 types: [fams 24 : types 24 : tokens 24]

[achieve_\[1\]](#) [aspect_\[1\]](#) [category_\[1\]](#) [complex_\[1\]](#) [constitution_\[1\]](#)
[construct_\[1\]](#) [define_\[1\]](#) [despite_\[1\]](#) [distinct_\[1\]](#) [dominate_\[1\]](#) [ensure_\[1\]](#)
[estimate_\[1\]](#) [function_\[1\]](#) [integrate_\[1\]](#) [intelligence_\[1\]](#) [interact_\[1\]](#) [media_\[1\]](#)
[migrate_\[1\]](#) [minimum_\[1\]](#) [participate_\[1\]](#) [radical_\[1\]](#) [source_\[1\]](#) [sum_\[1\]](#)
[unique_\[1\]](#)

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

[contrary_\[1\]](#) [predominant_\[1\]](#)

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams 1 : types 1 : tokens 1]

[consonant_\[1\]](#)

7.4.2.2 Indigenous Peoples

Fams	Freq	Range	BNCoca	Found in these texts			
remove	53	3	2		T_2	T_3	T_4
environment	27	3	2		T_2	T_3	T_4
area	13	4	1	T_1	T_2	T_3	T_4
adapt	13	1	2		T_2		
tradition	12	2	2		T_2	T_3	
establish	10	4	2	T_1	T_2	T_3	T_4

region	10	2	2 T_1	T_2		
policy	9	2	2	T_2	T_4	
create	8	2	2 T_1	T_2		
identify	8	2	2	T_2	T_3	
economy	8	1	2	T_2		
culture	7	2	2	T_2		T_4
style	7	2	2	T_2		T_4
community	6	2	2	T_2	T_3	
define	6	2	3	T_2	T_3	
final	6	2	1	T_2	T_3	
resource	6	2	3	T_2	T_3	
estimate	5	4	3 T_1	T_2	T_3	T_4
debate	5	2	3	T_2		T_4
revolution	5	2	3 T_1	T_2		
source	5	2	3	T_2	T_3	
domesticate	5	1	8	T_2		
federal	5	1	3	T_2		
persist	5	1	3	T_2		
text	5	1	3	T_2		
ethnic	4	2	3	T_2	T_3	
primary	4	2	3	T_2		T_4
migrate	4	1	3	T_2		
conflict	4	1	3	T_2		
impact	4	1	3	T_2		
major	4	1	1	T_2		
process	4	1	2	T_2		
similar	4	1	2	T_2		
authority	3	2	3	T_2		T_4
generation	3	2	2	T_2		T_4
perspective	3	2	3	T_2	T_3	
seek	3	2	2 T_1	T_2		
assist	3	1	2	T_2		
dominate	3	1	3	T_2		
factor	3	1	3	T_2		
occupy	3	1	3	T_2		
prime	3	1	2			T_4
range	3	1	2	T_2		
role	3	1	2	T_2		
immigrate	2	3	4 T_1	T_2	T_3	
circumstance	2	2	2	T_2		T_4
construct	2	2	3 T_1	T_2		
income	2	2	2	T_2	T_3	
issue	2	2	1	T_2	T_3	
publish	2	2	3 T_1	T_2		
transport	2	2	3 T_1	T_2		
grant	2	1	2	T_2		

edit	2	1	2			T_2
aspect	2	1	3			T_2
available	2	1	2			T_2
challenge	2	1	2			T_2
commodity	2	1	4			T_2
complex	2	1	3			T_2
consist	2	1	3 T_1			
contact	2	1	2			T_2
contemporary	2	1	3			T_2
element	2	1	3			T_2
emerge	2	1	3			T_2
encounter	2	1	3			T_2
evident	2	1	3			T_2
exploit	2	1	3			T_2
guarantee	2	1	2			T_2
instruct	2	1	2			T_2
integrity	2	1	4			T_2
interpret	2	1	3			T_2
invest	2	1	3			T_2
journal	2	1	3			T_2
liberal	2	1	3			T_2
maintain	2	1	2			T_2
motive	2	1	3			T_2
period	2	1	2			T_2
physical	2	1	2			T_2
portion	2	1	3			T_2
pursue	2	1	3			T_2
require	2	1	2			T_2
significant	2	1	3			T_2
vary	2	1	2			T_2
voluntary	2	1	3			T_2
civil	1	1	3			T_2
acquire	1	1	3			T_2
adequate	1	1	3			T_2
affect	1	1	2			T_2
annual	1	1	3			T_2
apparent	1	1	1			T_2
appropriate	1	1	3			T_2
approximate	1	1	3			T_2
assign	1	1	3			T_2
assure	1	1	2			T_2
attach	1	1	2			T_2
attitude	1	1	2			T_2
behalf	1	1	4 T_1			
bulk	1	1	4			T_2
clarify	1	1	3			T_2

commission	1	1	3	T_2	
consequence	1	1	3		
constitute	1	1	3	T_2	
consult	1	1	3	T_2	
context	1	1	3	T_2	
definitive	1	1	5	T_2	
design	1	1	2	T_2	
despite	1	1	3	T_2	
distinct	1	1	3	T_2	
document	1	1	3	T_2	
domain	1	1	4	T_2	
dominant	1	1	3	T_1	
drama	1	1	2	T_2	
dynamic	1	1	4	T_2	
emphasise	1	1	3	T_2	
eventually	1	1	2	T_1	
expose	1	1	2	T_2	
file	1	1	2		
finance	1	1	2	T_2	
function	1	1	3	T_1	
fund	1	1	2	T_2	
ignore	1	1	2	T_2	
image	1	1	2	T_2	
implicate	1	1	3	T_2	
inherent	1	1	4	T_2	
institute	1	1	3	T_2	
integrate	1	1	3	T_2	
interact	1	1	3	T_2	
involve	1	1	1	T_2	
job	1	1	1		T_3
justify	1	1	3	T_2	
legislate	1	1	3		T_3
link	1	1	3	T_2	
locate	1	1	2	T_2	
margin	1	1	3	T_2	
mediate	1	1	4	T_2	
military	1	1	2	T_2	
minor	1	1	2		T_3
notion	1	1	3	T_2	
obvious	1	1	1	T_2	
occur	1	1	2	T_2	
parallel	1	1	3	T_2	
perceive	1	1	3		
percent	1	1	2		T_3
philosophy	1	1	3	T_2	
potential	1	1	3	T_2	

precede	1	1	3	T_2	
previous	1	1	2	T_2	
project	1	1	2		T_4
promote	1	1	3	T_2	
quote	1	1	2	T_2	
radical	1	1	3	T_2	
reside	1	1	3	T_2	
retain	1	1	3	T_2	
scheme	1	1	3	T_2	
section	1	1	2	T_2	
secure	1	1	1	T_2	
series	1	1	2	T_2	
specific	1	1	2	T_2	
stress	1	1	2	T_2	
structure	1	1	3	T_2	
submit	1	1	3	T_2	
sum	1	1	3	T_2	
supplement	1	1	3	T_2	
symbol	1	1	3	T_2	
technology	1	1	2	T_2	
ultimate	1	1	3	T_2	
undergo	1	1	3	T_2	
unify	1	1	4	T_2	
unique	1	1	3	T_2	
utilise	1	1	4	T_2	
visual	1	1	3	T_2	

BNC-COCA-1,000 types: [fams 5 : types 5 : tokens 5]

[apparent](#) [1] [involve](#) [1] [job](#) [1] [obvious](#) [1] [secure](#) [1]

BNC-COCA-2,000 types: [fams 26 : types 26 : tokens 26]

[affect](#) [1] [assure](#) [1] [attach](#) [1] [attitude](#) [1] [design](#) [1] [drama](#) [1] [eventually](#) [1] [expose](#) [1] [file](#) [1] [finance](#) [1] [fund](#) [1] [ignore](#) [1] [image](#) [1] [locate](#) [1] [military](#) [1] [minor](#) [1] [occur](#) [1] [percent](#) [1] [previous](#) [1] [project](#) [1] [quote](#) [1] [section](#) [1] [series](#) [1] [specific](#) [1] [stress](#) [1] [technology](#) [1]

BNC-COCA-3,000 types: [fams 47 : types 47 : tokens 47]

[acquire](#) [1] [adequate](#) [1] [annual](#) [1] [appropriate](#) [1] [approximate](#) [1] [assign](#) [1] [civil](#) [1] [clarify](#) [1] [commission](#) [1] [consequence](#) [1] [constitute](#) [1] [consult](#) [1] [context](#) [1] [despite](#) [1] [distinct](#) [1] [document](#) [1] [dominant](#) [1] [emphasise](#) [1] [function](#) [1] [implicate](#) [1] [institute](#) [1] [integrate](#) [1] [interact](#) [1] [justify](#) [1] [legislate](#) [1] [link](#) [1] [margin](#) [1]

notion_[1] parallel_[1] perceive_[1] philosophy_[1] potential_[1] precede_[1]
 promote_[1] radical_[1] reside_[1] retain_[1] scheme_[1] structure_[1]
 submit_[1] sum_[1] supplement_[1] symbol_[1] ultimate_[1] undergo_[1]
 unique_[1] visual_[1]

BNC-COCA-4,000 types: [fams 8 : types 8 : tokens 8]

behalf_[1] bulk_[1] domain_[1] dynamic_[1] inherent_[1] mediate_[1]
 unify_[1] utilise_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

definitive_[1]

7.4.3 Targets

7.4.3.1 Global English

fam	frequency	range	bncoca	texts found in			
globe	19	4	3	T_1	T_2	T_3	T_4
communicate	16	3	3	T_1	T_2		T_4
culture	14	3	2	T_1		T_3	T_4
economy	8	2	2			T_3	T_4
area	6	4	1	T_1	T_2	T_3	T_4
region	6	3	2	T_1	T_2	T_3	
technology	6	3	2		T_2	T_3	T_4
establish	6	2	2	T_1		T_3	
range	5	3	2	T_1	T_2		T_4
military	5	1	2			T_3	
estimate	4	4	3	T_1	T_2	T_3	T_4
text	4	3	3	T_1		T_3	T_4
percent	4	3	2	T_1	T_2	T_3	
primary	4	2	3		T_2	T_3	
instance	4	2	2	T_1	T_2		
role	4	2	2		T_2	T_3	
computer	4	2	1		T_2		T_4
dominant	4	1	3		T_2		
feature	4	1	2	T_1			
challenge	3	3	2	T_1		T_3	T_4
identify	3	3	2	T_1		T_3	T_4
major	3	3	1	T_1	T_2	T_3	
distinct	3	2	3	T_1			T_4
media	3	2	3		T_2	T_3	
vary	3	2	2	T_1	T_2		
immigrate	9	1	4			T_3	
revolution	3	1	3			T_3	
transport	3	1	3			T_3	

stress	3	1	2	T_1			
tradition	2	4	2	T_1	T_2	T_3	T_4
chart	2	2	3	T_1			T_4
conflict	2	2	3	T_1		T_3	
diverse	2	2	3	T_1	T_2		
institution	2	2	3		T_2		T_4
neutral	2	2	3	T_1		T_3	
route	2	2	3	T_1		T_3	
transform	2	2	3			T_3	T_4
contribute	2	2	2		T_2	T_3	
create	2	2	2	T_1			T_4
finance	2	2	2		T_2	T_3	
illustrate	2	2	2	T_1		T_3	
style	2	2	2			T_3	T_4
definite	2	2	1		T_2	T_3	
issue	2	2	1			T_3	T_4
secure	2	2	1			T_3	T_4
define	2	1	3		T_2		
distribute	2	1	3		T_2		
dominate	2	1	3			T_3	
expand	2	1	3			T_3	
unique	2	1	3	T_1			
visual	2	1	3				T_4
seek	2	1	2				T_4
site	2	1	2				T_4
version	2	1	2	T_1			
final	2	1	1			T_3	
entity	1	3	4	T_1	T_2		T_4
ratio	1	3	3	T_1	T_2	T_3	
brevity	1	1	11		T_2		
globe	1	1	5		T_2		
identical	1	1	4	T_1			
academy	1	1	3		T_2		
administration	1	1	3			T_3	
analyse	1	1	3				T_4
aspect	1	1	3			T_3	
author	1	1	3				T_4
category	1	1	3		T_2		
clarify	1	1	3		T_2		
complex	1	1	3			T_3	
component	1	1	3				T_4
conceive	1	1	3	T_1			
consequent	1	1	3			T_3	
considerable	1	1	3	T_1			
consist	1	1	3	T_1			

consult	1	1	3		T_4
core	1	1	3	T_1	
criteria	1	1	3	T_2	
despite	1	1	3		T_4
element	1	1	3		T_4
emerge	1	1	3	T_1	
enable	1	1	3	T_2	
facilitate	1	1	3	T_2	
factor	1	1	3		T_3
formula	1	1	3		T_4
founded	1	1	3		T_3
function	1	1	3	T_1	
furthermore	1	1	3	T_2	
ideal	1	1	3		T_3
interact	1	1	3		T_4
interpret	1	1	3	T_1	
justify	1	1	3	T_2	
label	1	1	3	T_2	
method	1	1	3		T_4
monitor	1	1	3		T_4
motive	1	1	3		T_4
mutual	1	1	3	T_2	
nevertheless	1	1	3	T_2	
notion	1	1	3	T_1	
occupy	1	1	3		T_3
outcome	1	1	3		T_4
precise	1	1	3	T_2	
predict	1	1	3	T_1	
promote	1	1	3	T_2	
prospect	1	1	3	T_2	
psychology	1	1	3		T_4
publish	1	1	3		T_4
revise	1	1	3	T_1	
source	1	1	3		T_4
statistic	1	1	3	T_2	
strategy	1	1	3		T_4
target	1	1	3	T_1	
task	1	1	3	T_1	
trend	1	1	3		T_4
visible	1	1	3		T_3
access	1	1	2		T_4
adult	1	1	2		T_4
appreciate	1	1	2	T_1	
benefit	1	1	2		T_3
enormous	1	1	2		T_3
expose	1	1	2		T_3

generation	1	1	2	T_1	
goal	1	1	2	T_1	
individual	1	1	2	T_1	
legal	1	1	2		T_3
maintain	1	1	2		T_3
minor	1	1	2		T_4
period	1	1	2		T_3
presume	1	1	2	T_2	
previous	1	1	2	T_2	
process	1	1	2		T_3
react	1	1	2	T_1	
rely	1	1	2		T_4
research	1	1	2		T_4
series	1	1	2		T_3
shift	1	1	2		T_3
trace	1	1	2	T_1	
whereas	1	1	2	T_1	
involve	1	1	1		T_4
job	1	1	1	T_2	
obvious	1	1	1	T_1	

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

[involve_\[1\]](#) [job_\[1\]](#) [obvious_\[1\]](#)

BNC-COCA-2,000 types: [fams 23 : types 23 : tokens 23]

[access_\[1\]](#) [adult_\[1\]](#) [appreciate_\[1\]](#) [benefit_\[1\]](#) [enormous_\[1\]](#) [expose_\[1\]](#)
[generation_\[1\]](#) [goal_\[1\]](#) [individual_\[1\]](#) [legal_\[1\]](#) [maintain_\[1\]](#) [minor_\[1\]](#)
[period_\[1\]](#) [presume_\[1\]](#) [previous_\[1\]](#) [process_\[1\]](#) [react_\[1\]](#) [rely_\[1\]](#)
[research_\[1\]](#) [series_\[1\]](#) [shift_\[1\]](#) [trace_\[1\]](#) [whereas_\[1\]](#)

BNC-COCA-3,000 types: [fams 54 : types 54 : tokens 54]

[academy_\[1\]](#) [administration_\[1\]](#) [analyse_\[1\]](#) [aspect_\[1\]](#) [author_\[1\]](#)
[category_\[1\]](#) [clarify_\[1\]](#) [complex_\[1\]](#) [component_\[1\]](#) [conceive_\[1\]](#)
[consequent_\[1\]](#) [considerable_\[1\]](#) [consist_\[1\]](#) [consult_\[1\]](#) [core_\[1\]](#) [criteria_\[1\]](#)
[despite_\[1\]](#) [element_\[1\]](#) [emerge_\[1\]](#) [enable_\[1\]](#) [facilitate_\[1\]](#) [factor_\[1\]](#)
[formula_\[1\]](#) [founded_\[1\]](#) [function_\[1\]](#) [furthermore_\[1\]](#) [ideal_\[1\]](#) [interact_\[1\]](#)
[interpret_\[1\]](#) [justify_\[1\]](#) [label_\[1\]](#) [method_\[1\]](#) [monitor_\[1\]](#) [motive_\[1\]](#)
[mutual_\[1\]](#) [nevertheless_\[1\]](#) [notion_\[1\]](#) [occupy_\[1\]](#) [outcome_\[1\]](#) [precise_\[1\]](#)
[predict_\[1\]](#) [promote_\[1\]](#) [prospect_\[1\]](#) [psychology_\[1\]](#) [publish_\[1\]](#) [ratio_\[1\]](#)
[revise_\[1\]](#) [source_\[1\]](#) [statistic_\[1\]](#) [strategy_\[1\]](#) [target_\[1\]](#) [task_\[1\]](#) [trend_\[1\]](#)
[visible_\[1\]](#)

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

entity_[1] identical_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

7.4.3.2 Indigenous Peoples

fams	frequency	range	BNCola	Found in these texts		
team	21	2	1	T_2	T_3	
culture	14	2	2	T_1	T_3	
percent	10	2	2	T_1	T_3	
area	7	3	1	T_1	T_2	T_3
federal	7	3	3	T_1	T_2	T_3
challenge	7	2	2	T_1	T_3	
media	6	1	3	T_3		
image	5	3	2	T_1	T_2	T_3
issue	5	3	1	T_1	T_2	T_3
advocate	5	2	3	T_2	T_3	
community	5	2	2	T_1	T_3	
environment	5	2	2	T_1	T_2	
region	5	2	2	T_1	T_3	
civil	5	1	3	T_3		
significant	4	2	3	T_2	T_3	
sustain	4	1	3	T_1		
individual	3	3	2	T_1	T_2	T_3
specific	3	3	2	T_1	T_2	T_3
contact	3	2	2	T_2	T_3	
create	3	2	2	T_1	T_2	
resource	3	2	3	T_1	T_2	
tradition	3	2	2	T_2	T_3	
unique	3	2	3	T_1	T_2	
ensure	3	1	3	T_1		
major	3	1	1	T_1		
negative	3	1	3		T_3	
resolution	3	1	3		T_3	
maintain	2	3	2	T_1	T_2	T_3
edit	2	2	2	T_2	T_3	
behalf	2	2	4	T_2	T_3	
confine	2	2	3	T_2	T_3	
contrast	2	2	3	T_1	T_2	
diverse	2	2	3	T_1	T_2	
factor	2	2	3	T_1	T_3	
fund	2	2	2	T_1	T_2	

imagine	2	2	1	T_1	T_2	
legal	2	2	2	T_1		T_3
promote	2	2	3		T_2	T_3
symbol	2	2	3		T_2	T_3
profession	2	1	2			T_3
site	2	1	2		T_2	
drama	2	1	2		T_2	
generation	2	1	2	T_1		
identify	2	1	2	T_1		
intense	2	1	2	T_1		
migrate	2	1	3		T_2	
period	2	1	2		T_2	
policy	2	1	2			T_3
secure	2	1	1	T_1		
comprehensive	1	2	3	T_1		T_3
corporate	1	2	3		T_2	T_3
tense	1	2	2	T_1	T_2	
via	1	2	3		T_2	T_3
incorporate	1	1	3			T_3
chapter	1	1	2			T_3
estimate	1	1	3	T_1		
academy	1	1	3	T_1		
accommodate	1	1	3	T_1		
adequate	1	1	3	T_1		
analyse	1	1	3			T_3
aspect	1	1	3		T_2	
attitude	1	1	2	T_1		
available	1	1	2	T_1		
brief	1	1	2			T_3
category	1	1	3		T_2	
chemical	1	1	3	T_1		
component	1	1	3			T_3
comprise	1	1	3	T_1		
consequent	1	1	3	T_1		
constitution	1	1	3	T_1		
constrain	1	1	3		T_2	
consumption	1	1	3	T_1		
contemporary	1	1	3		T_2	
context	1	1	3		T_2	
contract	1	1	2		T_2	
contrary	1	1	4			T_3
contribute	1	1	2			T_3
controversy	1	1	3	T_1		
convince	1	1	2		T_2	
cycle	1	1	3	T_1		

decade	1	1	3		T_3
define	1	1	3		T_3
disproportion	1	1	6	T_2	
derive	1	1	3		T_3
document	1	1	3		T_3
dominate	1	1	3	T_2	
economy	1	1	2	T_1	
energy	1	1	2	T_1	
establish	1	1	2		T_3
exclusive	1	1	3	T_2	
expansion	1	1	3	T_2	
exploit	1	1	3	T_1	
export	1	1	3	T_1	
facility	1	1	3	T_1	
feature	1	1	2	T_2	
file	1	1	2		T_3
final	1	1	1	T_1	
finance	1	1	2		T_3
focus	1	1	3	T_2	
foundation	1	1	3		T_3
function	1	1	3	T_2	
global	1	1	3	T_1	
guarantee	1	1	2	T_2	
ignore	1	1	2	T_1	
illustrate	1	1	2		T_3
indicate	1	1	2		T_3
infrastructure	1	1	4	T_2	
institute	1	1	3		T_3
integrate	1	1	3		T_3
integrity	1	1	4	T_1	
invest	1	1	3	T_1	
job	1	1	1	T_1	
military	1	1	2	T_2	
minor	1	1	2	T_2	
norm	1	1	4	T_2	
occupation	1	1	3	T_2	
occur	1	1	2	T_2	
participate	1	1	3	T_1	
persist	1	1	3	T_2	
pose	1	1	3		T_3
positive	1	1	2		T_3
predominant	1	1	4	T_1	
primary	1	1	3	T_1	
prior	1	1	3	T_2	
priority	1	1	3	T_2	
process	1	1	2	T_2	

psychology	1	1	3		T_3
range	1	1	2	T_1	
register	1	1	2		T_3
regulate	1	1	3		T_3
release	1	1	2	T_1	
remove	1	1	2		T_3
require	1	1	2	T_2	
research	1	1	2		T_3
response	1	1	3	T_1	
retain	1	1	3		T_3
role	1	1	2	T_2	
scope	1	1	3	T_1	
sector	1	1	3	T_1	
select	1	1	2		T_3
status	1	1	3	T_2	
style	1	1	2	T_1	
submit	1	1	3		T_3
technical	1	1	3	T_1	
terminate	1	1	4		T_3
trend	1	1	3		T_3
version	1	1	2		T_3
visible	1	1	3		T_3

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]

[final_\[1\]](#) [job_\[1\]](#)

BNC-COCA-2,000 types: [fams 33 : types 33 : tokens 33]

[attitude_\[1\]](#) [available_\[1\]](#) [brief_\[1\]](#) [chapter_\[1\]](#) [contract_\[1\]](#) [contribute_\[1\]](#)
[convince_\[1\]](#) [economy_\[1\]](#) [energy_\[1\]](#) [establish_\[1\]](#) [feature_\[1\]](#) [file_\[1\]](#)
[finance_\[1\]](#) [guarantee_\[1\]](#) [ignore_\[1\]](#) [illustrate_\[1\]](#) [indicate_\[1\]](#) [military_\[1\]](#)
[minor_\[1\]](#) [occur_\[1\]](#) [positive_\[1\]](#) [process_\[1\]](#) [range_\[1\]](#) [register_\[1\]](#)
[release_\[1\]](#) [remove_\[1\]](#) [require_\[1\]](#) [research_\[1\]](#) [role_\[1\]](#) [select_\[1\]](#) [style_\[1\]](#)
[tense_\[1\]](#) [version_\[1\]](#)

BNC-COCA-3,000 types: [fams 57 : types 57 : tokens 57]

[academy_\[1\]](#) [accommodate_\[1\]](#) [adequate_\[1\]](#) [analyse_\[1\]](#) [aspect_\[1\]](#)
[category_\[1\]](#) [chemical_\[1\]](#) [component_\[1\]](#) [comprehensive_\[1\]](#) [comprise_\[1\]](#)
[consequent_\[1\]](#) [constitution_\[1\]](#) [constrain_\[1\]](#) [consumption_\[1\]](#)
[contemporary_\[1\]](#) [context_\[1\]](#) [controversy_\[1\]](#) [corporate_\[1\]](#) [cycle_\[1\]](#)
[decade_\[1\]](#) [define_\[1\]](#) [derive_\[1\]](#) [document_\[1\]](#) [dominate_\[1\]](#) [estimate_\[1\]](#)
[exclusive_\[1\]](#) [expansion_\[1\]](#) [exploit_\[1\]](#) [export_\[1\]](#) [facility_\[1\]](#) [focus_\[1\]](#)
[foundation_\[1\]](#) [function_\[1\]](#) [global_\[1\]](#) [incorporate_\[1\]](#) [institute_\[1\]](#)
[integrate_\[1\]](#) [invest_\[1\]](#) [occupation_\[1\]](#) [participate_\[1\]](#) [persist_\[1\]](#) [pose_\[1\]](#)
[primary_\[1\]](#) [prior_\[1\]](#) [priority_\[1\]](#) [psychology_\[1\]](#) [regulate_\[1\]](#) [response_\[1\]](#)

retain [1] scope [1] sector [1] status [1] submit [1] technical [1] trend [1]
via [1] visible [1]

BNC-COCA-4,000 types: [fams 6 : types 6 : tokens 6]

contrary [1] infrastructure [1] integrity [1] norm [1] predominant [1]
terminate [1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

disproportion [1]

7.5 AWL in-text once

7.5.1 Textbook: Access to English

7.5.1.1 Tailored text: *Divided by a Common Language*

area [1] aware [1] communicate [1] context [1] culture [1] distinct [1]
economy [1] element [1] feature [1] final [1] generation [1] logic [1]
norm [1] text [1]

7.5.1.2 Tailored text: *A Global Language*

administer [1] alternative [1] aspect [1] category [1] communicate [1]
contrary [1] domestic [1] drama [1] error [1] export [1] ignorant [1]
immigrate [1] impact [1] media [1] minor [1] option [1] period [1]
predict [1] proportion [1] rely [1] reveal [1] role [1] strategy [1]
structure [1] technology [1] text [1] vary [1]

7.5.1.3 Authentic text: *Renaming English Brisbane Times*

academy [1] acknowledge [1] benefit [1] consult [1] context [1]
convene [1] diverse [1] emphasis [1] evolve [1] fund [1] generation [1]
interact [1] lecture [1] link [1] participate [1] perspective [1] precede [1]
region [1] relevant [1] site [1] source [1]

7.5.1.4 Tailored text: *Native Americans: Original Inhabitants*

access [1] approximate [1] benefit [1] community [1] concentrate [1]
contribute [1] culture [1] decline [1] displace [1] diverse [1]
environment [1] establish [1] expand [1] final [1] fundamental [1]
migrate [1] military [1] occur [1] priority [1] radical [1] region [1]
remove [1] source [1] structure [1] team [1] trace [1] tradition [1]

7.5.1.5 Tailored text: *Aboriginal Australians*

access [1] alternative [1] apparent [1] approximate [1] chart [1]
community [1] concept [1] contact [1] coordinate [1] create [1] culture [1]
debate [1] discriminate [1] domestic [1] economy [1] emerge [1]
enormous [1] establish [1] estimate [1] exclude [1] exhibit [1] grant [1]
identify [1] individual [1] involve [1] job [1] labour [1] link [1] obvious [1]
period [1] phenomenon [1] physical [1] precise [1] process [1]
remove [1] source [1] status [1] technique [1] technology [1] vary [1]

7.5.1.6 Tailored text: *Stolen Children*

aspect [1] federal [1] immigrate [1] issue [1] media [1] react [1]

7.5.1.7 Authentic text: *Native Americans In Business*

administer [1] attain [1] benefit [1] consist [1] create [1] enhance [1]
focus [1] goal [1] impact [1] innovate [1] major [1] maximise [1]
potential [1] range [1] stable [1] statistic [1] tradition [1]

7.5.2 Textbook: Stunt

7.5.2.1 Tailored text: *British vs. American English*

area [1] aspect [1] category [1] compute [1] distinct [1] evident [1]
instance [1] intelligence [1] stress [1] sum [1] tense [1] vary [1]

7.5.2.2 Tailored text: *English as a World Language*

constant [1] couple [1] culture [1] dominate [1] ensure [1] establish [1]
interact [1] isolate [1] job [1] major [1] minimum [1] participate [1]
require [1] role [1] status [1] vary [1]

7.5.2.3 Authentic text: *There Is an Epidemic UNESCO*

achieve [1] aid [1] circumstance [1] complex [1] constant [1]
constitute [1] construct [1] contact [1] contrary [1] culture [1] define [1]
despite [1] estimate [1] function [1] gender [1] identify [1] image [1]
individual [1] integrate [1] involve [1] media [1] migrate [1] minor [1]
policy [1] predominant [1] prime [1] radical [1] unique [1] vary [1]

7.5.2.4 Tailored text: *Native Americans*

alter [1] community [1] define [1] environment [1] establish [1]
estimate [1] file [1] identify [1] income [1] issue [1] minor [1] percent [1]
perspective [1] policy [1] primary [1] project [1] resource [1] source [1]
tradition [1]

7.5.2.5 Tailored text: *Australia: The Birth of a Nation*

behalf [1] construct [1] create [1] dominate [1] estimate [1] eventual [1]
function [1] publish [1] region [1] revolution [1] seek [1] site [1]
transport [1]

7.5.2.6 Tailored text: *Stolen Generation*

area_[1] authority_[1] circumstance_[1] consequent_[1] environment_[1] establish_[1] estimate_[1] final_[1] job_[1] legislate_[1] perceive_[1]

7.5.2.7 Authentic text: *Effects of Removal*

acquire_[1] adequate_[1] affect_[1] annual_[1] apparent_[1] appropriate_[1] approximate_[1] assign_[1] assure_[1] attach_[1] attitude_[1] bulk_[1] circumstance_[1] civil_[1] clarify_[1] commission_[1] construct_[1] consult_[1] context_[1] definite_[1] design_[1] despite_[1] distinct_[1] document_[1] domain_[1] drama_[1] dynamic_[1] emphasis_[1] estimate_[1] expose_[1] finance_[1] fund_[1] generation_[1] ignorant_[1] image_[1] immigrate_[1] implicate_[1] income_[1] inherent_[1] integrate_[1] interact_[1] involve_[1] issue_[1] justify_[1] link_[1] locate_[1] margin_[1] military_[1] motivate_[1] motive_[1] notion_[1] obvious_[1] occur_[1] parallel_[1] philosophy_[1] potential_[1] precede_[1] previous_[1] promote_[1] publish_[1] quote_[1] radical_[1] reside_[1] retain_[1] scheme_[1] section_[1] secure_[1] series_[1] site_[1] specific_[1] stress_[1] structure_[1] submit_[1] sum_[1] supplement_[1] symbol_[1] technology_[1] transport_[1] ultimate_[1] undergo_[1] unify_[1] unique_[1] utilise_[1] visual_[1]

7.5.3 Textbook: **Targets**

7.5.3.1 Tailored text: *The Flavours of English*

appreciate_[1] area_[1] challenge_[1] chart_[1] conceive_[1] conflict_[1] consequent_[1] considerable_[1] consist_[1] core_[1] create_[1] diverse_[1] emerge_[1] expand_[1] function_[1] generation_[1] globe_[1] goal_[1] identical_[1] identify_[1] illustrate_[1] individual_[1] interpret_[1] major_[1] neutral_[1] notion_[1] obvious_[1] percent_[1] predict_[1] react_[1] revise_[1] route_[1] similar_[1] target_[1] task_[1] text_[1] trace_[1] tradition_[1] whereas_[1]

7.5.3.2 Tailored text: *The Power of English Part 1*

academy_[1] brief_[1] category_[1] clarify_[1] compute_[1] contribute_[1] criteria_[1] definite_[1] diverse_[1] enable_[1] entity_[1] facilitate_[1] finance_[1] furthermore_[1] instance_[1] institute_[1] job_[1] justify_[1] label_[1] mutual_[1] nevertheless_[1] precise_[1] presume_[1] previous_[1] promote_[1] prospect_[1] range_[1] ratio_[1] region_[1] role_[1] statistic_[1] technology_[1] vary_[1]

7.5.3.3 Tailored text: *The Power of English Part 2*

aspect_[1] benefit_[1] challenge_[1] complex_[1] conflict_[1] contribute_[1] definite_[1] enormous_[1] expose_[1] factor_[1] finance_[1] founded_[1] identify_[1] ideology_[1] illustrate_[1] issue_[1] legal_[1] maintain_[1] major_[1] media_[1] neutral_[1] occupy_[1] percent_[1] period_[1] process_[1] region_[1] route_[1] secure_[1] series_[1] shift_[1] style_[1] text_[1] transform_[1] visible_[1]

7.5.3.4 Authentic text: *English and the Future*

access_[1] adult_[1] analyse_[1] area_[1] author_[1] challenge_[1] chart_[1] component_[1] consequent_[1] consult_[1] create_[1] despite_[1] distinct_[1] economy_[1] element_[1] formula_[1] identify_[1] institute_[1] interact_[1] involve_[1] issue_[1] method_[1] minor_[1] monitor_[1] motivate_[1] outcome_[1] psychology_[1] publish_[1] rely_[1] research_[1] secure_[1] source_[1] specific_[1] strategy_[1] style_[1] text_[2] tradition_[1] transform_[1] trend_[1]

7.5.3.5 Tailored text: *Native Americans: We Are Still Here*

aspect_[1] behalf_[1] category_[1] confine_[1] consequent_[1] constrain_[1] contemporary_[1] context_[1] contract_[1] contrast_[1] convince_[1] corporate_[1] diverse_[1] dominate_[1] environment_[1] exclude_[1] expand_[1] feature_[1] focus_[1] function_[1] fund_[1] guarantee_[1] image_[1] individual_[1] infrastructure_[1] maintain_[1] military_[1] minor_[1] norm_[1] occupy_[1] occur_[1] persist_[1] prior_[1] priority_[1] process_[1] promote_[1] proportion_[1] require_[1] role_[1] specific_[1] status_[1] symbol_[1] tense_[1] unique_[1]

7.5.3.6 Tailored text: *Australia – the Island Continent*

academy_[1] accommodate_[1] adequate_[1] attitude_[1] available_[1] chemical_[1] comprise_[1] consequent_[1] constitute_[1] consume_[1] contrast_[1] controversy_[1] create_[1] cycle_[1] diverse_[1] economy_[1] energy_[1] estimate_[1] exploit_[1] export_[1] facilitate_[1] factor_[1] final_[1] fund_[1] globe_[1] ignorant_[1] image_[1] individual_[1] institute_[1] integrity_[1] intense_[2] invest_[1] issue_[1] job_[1] legal_[1] participate_[1] predominant_[1] primary_[1] range_[1] release_[1] resource_[1] respond_[1] scope_[1] sector_[1] specific_[1] style_[1] technical_[1]

7.5.3.7 Authentic text: *Ending the Era of Harmful “Indian” Mascots*

analyse_[1] area_[1] behalf_[1] brief_[1] chapter_[1] component_[1] comprehensive_[1] confine_[1] contact_[1] contrary_[1] contribute_[1] decade_[1] define_[1] derive_[1] document_[1] dominate_[1] edit_[1] establish_[1] factor_[1] federal_[1] file_[1] finance_[1] foundation_[1] illustrate_[1] incorporate_[1] indicate_[1] individual_[1] institute_[1]

integrate [1] issue [1] legal [1] maintain [1] pose [1] positive [1]
promote [1] psychology [1] region [1] regulate [1] remove [1]
research [1] retain [1] select [1] specific [1] submit [1] symbol [1]
terminate [1] tradition [1] trend [1] version [1] via [1] visible [1]

7.5.4 Text only file

area [1] aware [1] communicate [1] context [1] culture [1] distinct [1] economy [1]
element [1] feature [1] final [1] generation [1] logic [1] norm [1] text [1]

administer [1] alternative [1] aspect [1] category [1] communicate [1] contrary [1]
domestic [1] drama [1] error [1] export [1] final [2] ignorant [1] immigrate [1] impact [1]
media [1] minor [1] option [1] period [1] predict [1] proportion [1] rely [1] reveal [1]
role [1] strategy [1] structure [1] technology [1] text [1] vary [1]

academy [1] acknowledge [1] benefit [1] consult [1] context [1] convene [1] diverse [1]
emphasis [1] evolve [1] fund [1] generation [1] interact [1] lecture [1] link [1]
participate [1] perspective [1] precede [1] region [1] relevant [1] site [1] source [1]

access [1] approximate [1] benefit [1] community [1] concentrate [1] contribute [1]
culture [1] decline [1] displace [1] diverse [1] environment [1] establish [1] expand [1]
final [1] fundamental [1] migrate [1] military [1] occur [1] priority [1] radical [1]
region [1] remove [1] source [1] structure [1] team [1] trace [1] tradition [1]

access [1] alternative [1] apparent [1] approximate [1] chart [1] community [1]
concept [1] contact [1] coordinate [1] create [1] culture [1] debate [1] discriminate [1]
domestic [1] economy [1] emerge [1] enormous [1] establish [1] estimate [1] exclude [1]
exhibit [1] grant [1] identify [1] individual [1] involve [1] job [1] labour [1] link [1]
obvious [1] period [1] phenomenon [1] physical [1] precise [1] process [1] remove [1]
source [1] status [1] technique [1] technology [1] vary [1]

aspect [1] federal [1] immigrate [1] issue [1] media [1] react [1]

administer [1] attain [1] benefit [1] consist [1] create [1] enhance [1] focus [1] goal [1]
impact [1] innovate [1] major [1] maximise [1] potential [1] range [1] stable [1]
statistic [1] tradition [1]

area [1] aspect [1] category [1] compute [1] distinct [1] evident [1] instance [1]
intelligence [1] stress [1] sum [1] tense [1] vary [1]

constant [1] couple [1] culture [1] dominate [1] ensure [1] establish [1] interact [1]
isolate [1] job [1] major [1] minimum [1] participate [1] require [1] role [1] status [1]
vary [1]

achieve [1] aid [1] circumstance [1] complex [1] constant [1] constitute [1] construct [1]
contact [1] contrary [1] culture [1] define [1] despite [1] estimate [1] function [1]
gender [1] identify [1] image [1] individual [1] integrate [1] involve [1] media [1]
migrate [1] minor [1] policy [1] predominant [1] prime [1] radical [1] unique [1] vary [1]

alter [1] community [1] define [1] environment [1] establish [1] estimate [1] file [1]
identify [1] income [1] issue [1] minor [1] percent [1] perspective [1] policy [1]
primary [1] project [1] resource [1] source [1] tradition [1]

behalf [1] construct [1] create [1] dominate [1] estimate [1] eventual [1] function [1]
publish [1] region [1] revolution [1] seek [1] site [1] transport [1]

area [1] authority [1] circumstance [1] consequent [1] environment [1] establish [1]
estimate [1] final [1] job [1] legislate [1] perceive [1]

acquire [1] adequate [1] affect [1] annual [1] apparent [1] appropriate [1] approximate [1]
assign [1] assure [1] attach [1] attitude [1] bulk [1] circumstance [1] civil [1] clarify [1]
commission [1] construct [1] consult [1] context [1] definite [1] design [1] despite [1]
distinct [1] document [1] domain [1] drama [1] dynamic [1] emphasis [1] estimate [1]
expose [1] finance [1] fund [1] generation [1] ignorant [1] image [1] immigrate [1]
implicate [1] income [1] inherent [1] integrate [1] interact [1] involve [1] issue [1]
justify [1] link [1] locate [1] margin [1] military [1] motivate [1] motive [1] notion [1]
obvious [1] occur [1] parallel [1] philosophy [1] potential [1] precede [1] previous [1]
promote [1] publish [1] quote [1] radical [1] reside [1] retain [1] scheme [1] section [1]
secure [1] series [1] site [1] specific [1] stress [1] structure [1] submit [1] sum [1]
supplement [1] symbol [1] technology [1] transport [1] ultimate [1] undergo [1] unify [1]
unique [1] utilise [1] visual [1]

appreciate [1] area [1] challenge [1] chart [1] conceive [1] conflict [1] consequent [1]
considerable [1] consist [1] core [1] create [1] diverse [1] emerge [1] expand [1]
function [1] generation [1] globe [1] goal [1] identical [1] identify [1] illustrate [1]
individual [1] interpret [1] major [1] neutral [1] notion [1] obvious [1] percent [1]
predict [1] react [1] revise [1] route [1] similar [1] target [1] task [1] text [1] trace [1]
tradition [1] whereas [1]

academy [1] brief [1] category [1] clarify [1] compute [1] contribute [1] criteria [1]
definite [1] diverse [1] enable [1] entity [1] facilitate [1] finance [1] furthermore [1]
instance [1] institute [1] job [1] justify [1] label [1] mutual [1] nevertheless [1]
precise [1] presume [1] previous [1] promote [1] prospect [1] range [1] ratio [1]
region [1] role [1] statistic [1] technology [1] vary [1]

aspect [1] benefit [1] challenge [1] complex [1] conflict [1] contribute [1] definite [1]
enormous [1] expose [1] factor [1] finance [1] founded [1] identify [1] ideology [1]
illustrate [1] issue [1] legal [1] maintain [1] major [1] media [1] neutral [1] occupy [1]
percent [1] period [1] process [1] region [1] route [1] secure [1] series [1] shift [1]
style [1] text [1] transform [1] visible [1]

access [1] adult [1] analyse [1] area [1] author [1] challenge [1] chart [1] component [1]
consequent [1] consult [1] create [1] despite [1] distinct [1] economy [1] element [1]
formula [1] identify [1] institute [1] interact [1] involve [1] issue [1] method [1]
minor [1] monitor [1] motivate [1] outcome [1] psychology [1] publish [1] rely [1]
research [1] secure [1] source [1] specific [1] strategy [1] style [1] text [2] tradition [1]
transform [1] trend [1]

aspect [1] behalf [1] category [1] confine [1] consequent [1] constrain [1]
contemporary [1] context [1] contract [1] contrast [1] convince [1] corporate [1]
diverse [1] dominate [1] environment [1] exclude [1] expand [1] feature [1] focus [1]
function [1] fund [1] guarantee [1] image [1] individual [1] infrastructure [1] maintain [1]
military [1] minor [1] norm [1] occupy [1] occur [1] persist [1] prior [1] priority [1]

process [1] promote [1] proportion [1] require [1] role [1] specific [1] status [1]
symbol [1] tense [1] unique [1]

academy [1] accommodate [1] adequate [1] attitude [1] available [1] chemical [1]
comprise [1] consequent [1] constitute [1] consume [1] contrast [1] controversy [1]
create [1] cycle [1] diverse [1] economy [1] energy [1] estimate [1] exploit [1] export [1]
facilitate [1] factor [1] final [1] fund [1] globe [1] ignorant [1] image [1] individual [1]
institute [1] integrity [1] intense [2] invest [1] issue [1] job [1] legal [1] participate [1]
predominant [1] primary [1] range [1] release [1] resource [1] respond [1] scope [1]
sector [1] specific [1] style [1] technical [1]

analyse [1] area [1] behalf [1] brief [1] chapter [1] component [1] comprehensive [1]
confine [1] contact [1] contrary [1] contribute [1] decade [1] define [1] derive [1]
document [1] dominate [1] edit [1] establish [1] factor [1] federal [1] file [1] finance [1]
foundation [1] illustrate [1] incorporate [1] indicate [1] individual [1] institute [1]
integrate [1] issue [1] legal [1] maintain [1] pose [1] positive [1] promote [1]
psychology [1] region [1] regulate [1] remove [1] research [1] retain [1] select [1]
specific [1] submit [1] symbol [1] terminate [1] tradition [1] trend [1] version [1] via [1]
visible [1]

7.5.6 Stunt AWL once total textbook

BNC-COCA-1,000 types: [fams 12 : types 12 : tokens 16]

apparent [1] area [2] compute [1] couple [1] definite [1] final [1]
involve [2] issue [2] job [2] major [1] obvious [1] secure [1]

BNC-COCA-2,000 types: [fams 50 : types 50 : tokens 66]

affect [1] aid [1] alter [1] assure [1] attach [1] attitude [1]
circumstance [3] community [1] constant [2] contact [1] create [1]
culture [2] design [1] drama [1] environment [2] establish [3] eventual [1]
expose [1] file [1] finance [1] fund [1] generation [1] identify [2]
image [2] income [2] individual [1] instance [1] locate [1] military [1]
minor [2] occur [1] percent [1] policy [2] previous [1] prime [1] project [1]
quote [1] region [1] require [1] role [1] section [1] seek [1] series [1]
site [2] specific [1] stress [2] technology [1] tense [1] tradition [1]
vary [3]

BNC-COCA-3,000 types: [fams 72 : types 73 : tokens 91]

achieve [1] acquire [1] adequate [1] annual [1] appropriate [1]
approximate [1] aspect [1] assign [1] authority [1] category [1] civil [1]
clarify [1] commission [1] complex [1] consequent [1] constitute [1]
construct [3] consult [1] context [1] define [2] despite [2] distinct [2]
document [1] dominate [2] emphasis [1] ensure [1] estimate [5]

evident [1] function [2] gender [1] implicate [1] integrate [2]
intelligence [1] interact [2] isolate [1] justify [1] legislate [1] link [1]
margin [1] media [1] migrate [1] minimum [1] motivate [1] motive [1]
notion [1] parallel [1] participate [1] perceive [1] perspective [1]
philosophy [1] potential [1] precede [1] primary [1] promote [1]
publish [2] radical [2] reside [1] resource [1] retain [1] revolution [1]
scheme [1] source [1] status [1] structure [1] submit [1] sum [2]
supplement [1] symbol [1] transport [2] ultimate [1] undergo [1]
unique [2] visual [1]

BNC-COCA-4,000 types: [fams 10 : types 10 : tokens 10]

behalf [1] bulk [1] contrary [1] domain [1] dynamic [1] immigrate [1]
inherent [1] predominant [1] unify [1] utilise [1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant [1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.5.6 Targets

BNC-COCA-1,000 types: [fams 11 : types 11 : tokens 306]

area_[3] compute_[1] definite_[2] final_[1] involve_[1] issue_[4] job_[2]
major_[2] number_[287] obvious_[1] secure_[2]

BNC-COCA-2,000 types: [fams 68 : types 68 : tokens 106]

access_[1] adult_[1] appreciate_[1] attitude_[1] available_[1] benefit_[1]
brief_[2] challenge_[3] chapter_[1] contact_[1] contract_[1] contribute_[3]
convince_[1] create_[3] economy_[2] edit_[1] energy_[1] enormous_[1]
environment_[1] establish_[1] expose_[1] feature_[1] file_[1] finance_[3]
fund_[2] generation_[1] goal_[1] guarantee_[1] identify_[3] illustrate_[3]
image_[2] indicate_[1] individual_[4] instance_[1] intense_[1] legal_[3]
maintain_[3] military_[1] minor_[2] occur_[1] percent_[2] period_[1]
positive_[1] presume_[1] previous_[1] process_[2] range_[2] react_[1]
region_[3] release_[1] rely_[1] remove_[1] require_[1] research_[2] role_[2]
select_[1] series_[1] shift_[1] similar_[1] specific_[4] style_[3] technology_[1]
tense_[1] trace_[1] tradition_[3] vary_[1] version_[1] whereas_[1]

BNC-COCA-3,000 types: [fams 113 : types 113 : tokens 149]

academy_[2] accommodate_[1] adequate_[1] analyse_[2] aspect_[2]
author_[1] category_[2] chart_[2] chemical_[1] clarify_[1] complex_[1]
component_[2] comprehensive_[1] comprise_[1] conceive_[1] confine_[2]
conflict_[2] consequent_[4] considerable_[1] consist_[1] constitute_[1]
constrain_[1] consult_[1] consume_[1] contemporary_[1] context_[1]
contrast_[2] controversy_[1] core_[1] corporate_[1] criteria_[1] cycle_[1]
decade_[1] define_[1] derive_[1] despite_[1] distinct_[1] diverse_[4]
document_[1] dominate_[2] element_[1] emerge_[1] enable_[1]
estimate_[1] exclude_[1] expand_[2] exploit_[1] export_[1] facilitate_[2]
factor_[3] federal_[1] focus_[1] formula_[1] foundation_[1] founded_[1]

function_[2] furthermore_[1] ideology_[1] incorporate_[1] institute_[4]
integrate_[1] interact_[1] interpret_[1] invest_[1] justify_[1] label_[1]
media_[1] method_[1] monitor_[1] motivate_[1] mutual_[1] neutral_[2]
nevertheless_[1] notion_[1] occupy_[2] outcome_[1] participate_[1]
persist_[1] pose_[1] precise_[1] predict_[1] primary_[1] prior_[1] priority_[1]
promote_[3] proportion_[1] prospect_[1] psychology_[2] publish_[1] ratio_[1]
regulate_[1] resource_[1] respond_[1] retain_[1] revise_[1] route_[2]
scope_[1] sector_[1] source_[1] statistic_[1] status_[1] strategy_[1]
submit_[1] symbol_[2] target_[1] task_[1] technical_[1] text_[3]
transform_[2] trend_[2] unique_[1] via_[1] visible_[2]

BNC-COCA-4,000 types: [fams 9 : types 9 : tokens 10]

behalf_[2] contrary_[1] entity_[1] identical_[1] infrastructure_[1] integrity_[1]
norm_[1] predominant_[1] terminate_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 2]

globe_[2]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

7.6 VP-compleat analysis of all AWL vocabulary

BNC-COCA-1,000 types: [fams 14 : types 14 : tokens 47]

apparent_2] area_6] aware_1] compute_2] couple_1] definite_3]
final_5] involve_4] issue_7] job_5] major_4] obvious_3] secure_3]
team_1]

BNC-COCA-2,000 types: [fams 96 : types 96 : tokens 228]

access_3] adult_1] affect_1] aid_1] alter_1] appreciate_1] assure_1]
attach_1] attitude_2] available_1] benefit_4] brief_2] challenge_3]
chapter_1] circumstance_3] community_3] concentrate_1] constant_2]
contact_3] contract_1] contribute_4] convince_1] create_6] culture_5]
design_1] drama_2] economy_4] edit_1] energy_1] enormous_2]
environment_4] establish_6] eventual_1] expose_2] feature_2] file_2]
finance_4] fund_4] generation_4] goal_2] grant_1] guarantee_1]
identify_6] illustrate_3] image_4] income_2] indicate_1] individual_6]
instance_2] intense_1] labour_1] legal_3] locate_1] maintain_3]
military_3] minor_5] occur_3] option_1] percent_3] period_3]
physical_1] policy_2] positive_1] presume_1] previous_2] prime_1]
process_3] project_1] quote_1] range_3] react_2] region_6] release_1]
rely_2] remove_3] require_2] research_2] role_4] section_1] seek_1]
select_1] series_2] shift_1] similar_1] site_3] specific_5] stable_1]
stress_2] style_3] technology_4] tense_2] trace_2] tradition_6] vary_6]
version_1] whereas_1]

BNC-COCA-3,000 types: [fams 174 : types 175 : tokens 313]

academy_3] accommodate_1] achieve_1] acknowledge_1] acquire_1]
adequate_2] alternative_2] analyse_2] annual_1] appropriate_1]
approximate_3] aspect_5] assign_1] author_1] authority_1] category_4]
chart_3] chemical_1] civil_1] clarify_2] commission_1] communicate_2]
complex_2] component_2] comprehensive_1] comprise_1] conceive_1]
concept_1] confine_2] conflict_2] consequent_5] considerable_1]
consist_2] constitute_2] constrain_1] construct_3] consult_3]
consume_1] contemporary_1] context_4] contrast_2] controversy_1]
coordinate_1] core_1] corporate_1] criteria_1] cycle_1] debate_1]
decade_1] decline_1] define_3] derive_1] despite_3] discriminate_1]
distinct_4] diverse_6] document_2] domestic_2] dominate_4]
element_2] emerge_2] emphasis_2] enable_1] enhance_1] ensure_1]
error_1] estimate_7] evident_1] evolve_1] exclude_2] exhibit_1]
expand_3] exploit_1] export_2] facilitate_2] factor_3] federal_2]
focus_2] formula_1] foundation_1] founded_1] function_4]
fundamental_1] furthermore_1] gender_1] ideology_1] impact_2]
implicate_1] incorporate_1] innovate_1] institute_4] integrate_3]
intelligence_1] interact_4] interpret_1] invest_1] isolate_1] justify_2]
label_1] lecture_1] legislate_1] link_2] logic_1] margin_1] media_4]
method_1] migrate_2] minimum_1] monitor_1] motivate_2] motive_1]
mutual_1] neutral_2] nevertheless_1] notion_2] occupy_2] outcome_1]
parallel_1] participate_3] perceive_1] persist_1] perspective_2]
phenomenon_1] philosophy_1] pose_1] potential_2] precede_2]
precise_2] predict_2] primary_2] prior_1] priority_2] promote_4]
proportion_2] prospect_1] psychology_2] publish_3] radical_3] ratio_1]
regulate_1] relevant_1] reside_1] resource_2] respond_1] retain_2]
reveal_1] revise_1] revolution_1] route_2] scheme_1] scope_1]
sector_1] source_5] statistic_2] status_3] strategy_2] structure_3]
submit_2] sum_2] supplement_1] symbol_3] target_1] task_1]
technical_1] technique_1] text_5] transform_2] transport_2] trend_2]
ultimate_1] undergo_1] unique_3] via_1] visible_2] visual_1]

BNC-COCA-4,000 types: [fams 20 : types 20 : tokens 29]

administer_2] attain_1] behalf_3] bulk_1] contrary_3] displace_1]
domain_1] dynamic_1] entity_1] identical_1] immigrate_3]
infrastructure_1] inherent_1] integrity_1] maximise_1] norm_2]
predominant_2] terminate_1] unify_1] utilise_1]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 3]

convene_1] globe_2]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 3]

[ignorant_\[3\]](#)

BNC-COCA-7,000 types: [fams : types : tokens]

BNC-COCA-8,000 types: [fams : types : tokens]

BNC-COCA-9,000 types: [fams : types : tokens]

BNC-COCA-10,000 types: [fams : types : tokens]

BNC-COCA-11,000 types: [fams : types : tokens]

BNC-COCA-12,000 types: [fams : types : tokens]

BNC-COCA-13,000 types: [fams : types : tokens]

BNC-COCA-14,000 types: [fams : types : tokens]

BNC-COCA-15,000 types: [fams : types : tokens]

BNC-COCA-16,000 types: [fams : types : tokens]

BNC-COCA-17,000 types: [fams : types : tokens]

BNC-COCA-18,000 types: [fams : types : tokens]

BNC-COCA-19,000 types: [fams : types : tokens]

BNC-COCA-20,000 types: [fams : types : tokens]

BNC-COCA-21,000 types: [fams : types : tokens]

BNC-COCA-22,000 types: [fams : types : tokens]

BNC-COCA-23,000 types: [fams : types : tokens]

BNC-COCA-24,000 types: [fams : types : tokens]

BNC-COCA-25,000 types: [fams : types : tokens]

OFFLIST: [?: types 0 : tokens 0]

Textbook: Access to English

Topic: English as a Universal Language

Tailored text: Divided by a Common Language:

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]
[area_\[1\]](#) [aware_\[1\]](#) [final_\[1\]](#)

BNC-COCA-2,000 types: [fams 7 : types 7 : tokens 7]

[culture_\[1\]](#) [economy_\[1\]](#) [eventual_\[1\]](#) [feature_\[1\]](#) [generation_\[1\]](#) [labour_\[1\]](#)
[prime_\[1\]](#)

BNC-COCA-3,000 types: [fams 7 : types 7 : tokens 7]

[communicate_\[1\]](#) [context_\[1\]](#) [distinct_\[1\]](#) [element_\[1\]](#) [logic_\[1\]](#)
[revolution_\[1\]](#) [text_\[1\]](#)

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

[norm_\[1\]](#)

Tailored text: A Global Language (34 wf)

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]

[final_\[1\]](#) [major_\[1\]](#)

BNC-COCA-2,000 types: [fams 9 : types 9 : tokens 9]

[culture_\[1\]](#) [drama_\[1\]](#) [minor_\[1\]](#) [option_\[1\]](#) [period_\[1\]](#) [rely_\[1\]](#) [role_\[1\]](#)
[technology_\[1\]](#) [vary_\[1\]](#)

BNC-COCA-3,000 types: [fams 18 : types 18 : tokens 18]

[alternative_\[1\]](#) [aspect_\[1\]](#) [category_\[1\]](#) [communicate_\[1\]](#) [conflict_\[1\]](#)
[domestic_\[1\]](#) [dominate_\[1\]](#) [error_\[1\]](#) [expert_\[1\]](#) [export_\[1\]](#) [impact_\[1\]](#)
[media_\[1\]](#) [predict_\[1\]](#) [proportion_\[1\]](#) [reveal_\[1\]](#) [strategy_\[1\]](#) [structure_\[1\]](#)
[text_\[1\]](#)

BNC-COCA-4,000 types: [fams 3 : types 3 : tokens 3]

[administer_\[1\]](#) [contrary_\[1\]](#) [immigrate_\[1\]](#)

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

Authentic text: Renaming English Brisbane Times

BNC-COCA-1,000 types: [fams 0 : types 0 : tokens 0]

BNC-COCA-2,000 types: [fams 8 : types 8 : tokens 8]

benefit_[1] culture_[1] economy_[1] fund_[1] generation_[1] region_[1]
research_[1] site_[1]

BNC-COCA-3,000 types: [fams 19 : types 19 : tokens 19]

academy_[1] acknowledge_[1] communicate_[1] consult_[1] context_[1]
diverse_[1] emphasis_[1] evolve_[1] impact_[1] interact_[1] journal_[1]
lecture_[1] link_[1] participate_[1] perspective_[1] precede_[1] publish_[1]
relevant_[1] source_[1]

BNC-COCA-4,000 types: [fams : types : tokens]

BNC-COCA-5,000 types: [fams 2 : types 2 : tokens 2]

convene_[1] globe_[1]

Topic: Indigenous peoples

Tailored text: Native Americans: Original Inhabitants (36 word families)

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

area_[1] final_[1] team_[1]

BNC-COCA-2,000 types: [fams 19 : types 19 : tokens 19]

access_[1] benefit_[1] community_[1] concentrate_[1] contact_[1]
contribute_[1] culture_[1] environment_[1] establish_[1] feature_[1]
locate_[1] military_[1] occur_[1] region_[1] remove_[1] survive_[1]
technology_[1] trace_[1] tradition_[1]

BNC-COCA-3,000 types: [fams 13 : types 13 : tokens 13]

approximate_[1] conflict_[1] convert_[1] decline_[1] diverse_[1] estimate_[1]
expand_[1] fundamental_[1] migrate_[1] priority_[1] radical_[1] source_[1]
structure_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

displace_[1]

Tailored text: Aboriginal Australians (51 word families)

BNC-COCA-1,000 types: [fams 6 : types 6 : tokens 6]

apparent_[1] area_[1] involve_[1] issue_[1] job_[1] obvious_[1]

BNC-COCA-2,000 types: [fams 22 : types 22 : tokens 22]

access_[1] adapt_[1] community_[1] contact_[1] create_[1] culture_[1]
economy_[1] enormous_[1] establish_[1] grant_[1] identify_[1] individual_[1]
labour_[1] period_[1] physical_[1] policy_[1] process_[1] remove_[1]
style_[1] survive_[1] technology_[1] vary_[1]

BNC-COCA-3,000 types: [fams 22 : types 22 : tokens 22]

alternative_[1] approximate_[1] chart_[1] civil_[1] concept_[1] conflict_[1]
coordinate_[1] debate_[1] decline_[1] discriminate_[1] domestic_[1]
emerge_[1] estimate_[1] exclude_[1] exhibit_[1] isolate_[1] link_[1]
phenomenon_[1] precise_[1] source_[1] status_[1] technique_[1]

BNC-COCA-4,000 types: [fams : types : tokens]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

Tailored text: Stolen Children (8 word families)

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]

issue_[1]

BNC-COCA-2,000 types: [fams 3 : types 3 : tokens 3]

generation_[1] react_[1] remove_[1]

BNC-COCA-3,000 types: [fams 3 : types 3 : tokens 3]

aspect_[1] federal_[1] media_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

immigrate_[1]

Authentic text: Native Americans In Business

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]

major_[1]

BNC-COCA-2,000 types: [fams 12 : types 12 : tokens 12]

assist_[1] benefit_[1] community_[1] contribute_[1] create_[1] economy_[1]
finance_[1] goal_[1] individual_[1] range_[1] stable_[1] tradition_[1]

BNC-COCA-3,000 types: [fams 10 : types 10 : tokens 10]

consist_[1] corporate_[1] enhance_[1] focus_[1] impact_[1] innovate_[1]
invest_[1] potential_[1] resource_[1] statistic_[1]

BNC-COCA-4,000 types: [fams 3 : types 3 : tokens 3]

administer_[1] attain_[1] maximise_[1]

Textbook: Stunt

Topic: English as a Universal Language

Tailored text: British vs. American English

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]
area_[1] compute_[1]

BNC-COCA-2,000 types: [fams 5 : types 5 : tokens 5]

instance_[1] quote_[1] stress_[1] tense_[1] vary_[1]

BNC-COCA-3,000 types: [fams 10 : types 10 : tokens 10]

analyse_[1] aspect_[1] category_[1] consist_[1] distinct_[1] emphasis_[1]
evident_[1] intelligence_[1] source_[1] sum_[1]

Tailored text: English as a World Language

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]
couple_[1] job_[1] major_[1]

BNC-COCA-2,000 types: [fams 7 : types 7 : tokens 7]

constant_[1] culture_[1] establish_[1] expose_[1] require_[1] role_[1]
vary_[1]

BNC-COCA-3,000 types: [fams 8 : types 8 : tokens 8]

communicate_[1] dominate_[1] ensure_[1] interact_[1] isolate_[1]
minimum_[1] participate_[1] status_[1]

Authentic text: There Is an Epidemic UNESCO

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]
involve_[1]

BNC-COCA-2,000 types: [fams 19 : types 19 : tokens 19]

aid_[1] circumstance_[1] community_[1] constant_[1] contact_[1] create_[1]
culture_[1] generation_[1] identify_[1] image_[1] individual_[1] maintain_[1]
minor_[1] policy_[1] prime_[1] region_[1] specific_[1] survive_[1] vary_[1]

BNC-COCA-3,000 types: [fams 24 : types 24 : tokens 24]

achieve_[1] communicate_[1] complex_[1] concept_[1] constitute_[1]
construct_[1] context_[1] decade_[1] define_[1] despite_[1] estimate_[1]
factor_[1] function_[1] gender_[1] initiate_[1] institute_[1] integrate_[1]
isolate_[1] media_[1] migrate_[1] radical_[1] status_[1] trend_[1] unique_[1]

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

contrary_[1] predominant_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

Topic: Indigenous peoples

Tailored text: Native Americans

BNC-COCA-1,000 types: [fams 2 : types 2 : tokens 2]

area_[1] issue_[1]

BNC-COCA-2,000 types: [fams 13 : types 13 : tokens 13]

alter_[1] community_[1] environment_[1] establish_[1] file_[1] identify_[1]
income_[1] minor_[1] percent_[1] policy_[1] project_[1] remove_[1]
tradition_[1]

BNC-COCA-3,000 types: [fams 7 : types 7 : tokens 7]

define_[1] estimate_[1] ethnic_[1] perspective_[1] primary_[1] resource_[1]
source_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

immigrate_[1]

Tailored text: Australia: The Birth of a Nation

BNC-COCA-1,000 types: [fams 1 : types 1 : tokens 1]

area_[1]

BNC-COCA-2,000 types: [fams 6 : types 6 : tokens 6]

create_[1] establish_[1] eventual_[1] region_[1] seek_[1] site_[1]

BNC-COCA-3,000 types: [fams 8 : types 8 : tokens 8]

consist_[1] construct_[1] dominate_[1] estimate_[1] function_[1] publish_[1]
revolution_[1] transport_[1]

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

behalf_[1] immigrate_[1]

Tailored text: Stolen Generation

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

area_[1] final_[1] job_[1]

BNC-COCA-2,000 types: [fams 8 : types 8 : tokens 8]

circumstance_[1] culture_[1] environment_[1] establish_[1] generation_[1]
prime_[1] remove_[1] style_[1]

BNC-COCA-3,000 types: [fams 6 : types 6 : tokens 6]

authority_[1] consequent_[1] debate_[1] estimate_[1] legislate_[1]
perceive_[1]

Authentic text: Effects of Removal

BNC-COCA-1,000 types: [fams 9 : types 9 : tokens 9]

apparent_[1] area_[1] definite_[1] final_[1] involve_[1] issue_[1] major_[1]
obvious_[1] secure_[1]

BNC-COCA-2,000 types: [fams 55 : types 55 : tokens 55]

adapt_[1] affect_[1] assist_[1] assure_[1] attach_[1] attitude_[1]
available_[1] challenge_[1] circumstance_[1] community_[1] contact_[1]
create_[1] culture_[1] design_[1] drama_[1] economy_[1] edit_[1]
environment_[1] establish_[1] expose_[1] finance_[1] fund_[1]
generation_[1] grant_[1] guarantee_[1] identify_[1] image_[1] income_[1]
instruct_[1] locate_[1] maintain_[1] military_[1] occur_[1] period_[1]
physical_[1] policy_[1] previous_[1] process_[1] quote_[1] range_[1]
region_[1] remove_[1] require_[1] role_[1] section_[1] seek_[1] series_[1]
similar_[1] site_[1] specific_[1] stress_[1] style_[1] technology_[1]
tradition_[1] vary_[1]

BNC-COCA-3,000 types: [fams 83 : types 83 : tokens 83]

acquire_[1] adequate_[1] annual_[1] appropriate_[1] approximate_[1]
aspect_[1] assign_[1] authority_[1] civil_[1] clarify_[1] commission_[1]
complex_[1] conflict_[1] constitute_[1] construct_[1] consult_[1]
contemporary_[1] context_[1] debate_[1] define_[1] despite_[1] distinct_[1]
document_[1] domestic_[1] dominate_[1] element_[1] emerge_[1]
emphasis_[1] encounter_[1] estimate_[1] ethnic_[1] evident_[1] exploit_[1]
factor_[1] federal_[1] impact_[1] implicate_[1] institute_[1] integrate_[1]
interact_[1] interpret_[1] invest_[1] journal_[1] justify_[1] liberal_[1] link_[1]
margin_[1] migrate_[1] motivate_[1] motive_[1] notion_[1] occupy_[1]
parallel_[1] persist_[1] perspective_[1] philosophy_[1] portion_[1]
potential_[1] precede_[1] primary_[1] promote_[1] publish_[1] pursue_[1]
radical_[1] reside_[1] resource_[1] retain_[1] revolution_[1] scheme_[1]
significant_[1] source_[1] structure_[1] submit_[1] sum_[1] supplement_[1]
symbol_[1] text_[1] transport_[1] ultimate_[1] undergo_[1] unique_[1]
visual_[1] voluntary_[1]

BNC-COCA-4,000 types: [fams 9 : types 9 : tokens 9]

bulk_[1] commodity_[1] domain_[1] dynamic_[1] immigrate_[1] inherent_[1]
integrity_[1] unify_[1] utilise_[1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

Textbook: Targets

Topic: English as a Universal Language

Tailored text: The Flavours of English

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

area_[1] major_[1] obvious_[1]

BNC-COCA-2,000 types: [fams 23 : types 23 : tokens 23]

appreciate_[1] challenge_[1] create_[1] culture_[1] establish_[1] feature_[1]
generation_[1] goal_[1] identify_[1] illustrate_[1] individual_[1] instance_[1]
percent_[1] range_[1] react_[1] region_[1] similar_[1] stress_[1] trace_[1]
tradition_[1] vary_[1] version_[1] whereas_[1]

BNC-COCA-3,000 types: [fams 23 : types 23 : tokens 23]

chart_[1] communicate_[1] conceive_[1] conflict_[1] consequent_[1]
considerable_[1] consist_[1] core_[1] distinct_[1] diverse_[1] emerge_[1]
expand_[1] function_[1] interpret_[1] neutral_[1] notion_[1] predict_[1]
revise_[1] route_[1] target_[1] task_[1] text_[1] unique_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

identical_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

Tailored text: The Power of English Part 1

BNC-COCA-1,000 types: [fams 5 : types 5 : tokens 5]

area_[1] compute_[1] definite_[1] job_[1] major_[1]

BNC-COCA-2,000 types: [fams 12 : types 12 : tokens 12]

brief_[1] contribute_[1] finance_[1] instance_[1] percent_[1] presume_[1]
previous_[1] range_[1] region_[1] role_[1] technology_[1] vary_[1]

BNC-COCA-3,000 types: [fams 25 : types 25 : tokens 25]

academy_[1] category_[1] clarify_[1] communicate_[1] criteria_[1] define_[1]
distribute_[1] diverse_[1] dominate_[1] enable_[1] estimate_[1] facilitate_[1]
furthermore_[1] institute_[1] justify_[1] label_[1] media_[1] mutual_[1]
nevertheless_[1] precise_[1] primary_[1] promote_[1] prospect_[1] ratio_[1]
statistic_[1]

BNC-COCA-4,000 types: [fams 1 : types 1 : tokens 1]

entity_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

Tailored text: The Power of English Part 2

BNC-COCA-1,000 types: [fams 6 : types 6 : tokens 6]

area_[1] definite_[1] final_[1] issue_[1] major_[1] secure_[1]

BNC-COCA-2,000 types: [fams 23 : types 23 : tokens 23]

benefit_[1] challenge_[1] contribute_[1] culture_[1] economy_[1]
enormous_[1] establish_[1] expose_[1] finance_[1] identify_[1] illustrate_[1]
legal_[1] maintain_[1] military_[1] percent_[1] period_[1] process_[1]
region_[1] role_[1] series_[1] shift_[1] style_[1] technology_[1]

BNC-COCA-3,000 types: [fams 19 : types 19 : tokens 19]

aspect_[1] complex_[1] conflict_[1] consequent_[1] dominate_[1]
expand_[1] factor_[1] founded_[1] ideology_[1] media_[1] neutral_[1]
occupy_[1] primary_[1] revolution_[1] route_[1] text_[1] transform_[1]
transport_[1] visible_[1]

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

administer_[1] immigrate_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

Authentic text: English and the Future BBC

BNC-COCA-1,000 types: [fams 5 : types 5 : tokens 5]

area_[1] compute_[1] involve_[1] issue_[1] secure_[1]

BNC-COCA-2,000 types: [fams 17 : types 17 : tokens 17]

access_[1] adult_[1] challenge_[1] create_[1] culture_[1] economy_[1]
identify_[1] minor_[1] range_[1] rely_[1] research_[1] seek_[1] site_[1]
specific_[1] style_[1] technology_[1] tradition_[1]

BNC-COCA-3,000 types: [fams 25 : types 25 : tokens 25]

analyse_[1] author_[1] chart_[1] communicate_[1] component_[1]
consequent_[1] consult_[1] despite_[1] distinct_[1] element_[1] formula_[1]
institute_[1] interact_[1] method_[1] monitor_[1] motivate_[1] outcome_[1]
psychology_[1] publish_[1] source_[1] strategy_[1] text_[1] transform_[1]
trend_[1] visual_[1]

BNC-COCA-4,000 types: [fams : types : tokens]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

Topic: Indigenous peoples

Tailored text: Native Americans: We Are Still Here

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

area_[1] issue_[1] team_[1]

BNC-COCA-2,000 types: [fams 23 : types 23 : tokens 23]

contact_[1] contract_[1] convince_[1] create_[1] drama_[1] environment_[1]
feature_[1] fund_[1] guarantee_[1] image_[1] individual_[1] maintain_[1]
military_[1] minor_[1] occur_[1] period_[1] process_[1] require_[1] role_[1]
site_[1] specific_[1] tense_[1] tradition_[1]

BNC-COCA-3,000 types: [fams 29 : types 29 : tokens 29]

advocate_[1] aspect_[1] category_[1] confine_[1] consequent_[1]
constrain_[1] contemporary_[1] context_[1] contrast_[1] corporate_[1]
diverse_[1] dominate_[1] exclude_[1] expand_[1] federal_[1] focus_[1]
function_[1] migrate_[1] occupy_[1] persist_[1] prior_[1] priority_[1]
promote_[1] proportion_[1] resource_[1] significant_[1] status_[1]
symbol_[1] unique_[1]

BNC-COCA-4,000 types: [fams 3 : types 3 : tokens 3]

behalf_[1] infrastructure_[1] norm_[1]

Tailored text: Australia – the Island Continent

BNC-COCA-1,000 types: [fams 6 : types 6 : tokens 6]

area_[1] final_[1] issue_[1] job_[1] major_[1] secure_[1]

BNC-COCA-2,000 types: [fams 23 : types 23 : tokens 23]

attitude_[1] available_[1] challenge_[1] community_[1] create_[1] culture_[1]
economy_[1] energy_[1] environment_[1] fund_[1] generation_[1]
identify_[1] image_[1] individual_[1] intense_[1] legal_[1] maintain_[1]
percent_[1] range_[1] region_[1] release_[1] specific_[1] style_[1]

BNC-COCA-3,000 types: [fams 31 : types 31 : tokens 31]

academy_[1] accommodate_[1] adequate_[1] chemical_[1] comprise_[1]
consequent_[1] constitute_[1] consume_[1] contrast_[1] controversy_[1]
cycle_[1] diverse_[1] ensure_[1] estimate_[1] exploit_[1] export_[1]
facilitate_[1] factor_[1] federal_[1] institute_[1] invest_[1] migrate_[1]
participate_[1] primary_[1] resource_[1] respond_[1] scope_[1] sector_[1]
sustain_[1] technical_[1] unique_[1]

BNC-COCA-4,000 types: [fams 2 : types 2 : tokens 2]

integrity_[1] predominant_[1]

BNC-COCA-5,000 types: [fams 1 : types 1 : tokens 1]

globe_[1]

BNC-COCA-6,000 types: [fams 1 : types 1 : tokens 1]

ignorant_[1]

Indian Mascots

BNC-COCA-1,000 types: [fams 3 : types 3 : tokens 3]

area_[1] issue_[1] team_[1]

BNC-COCA-2,000 types: [fams 29 : types 29 : tokens 29]

brief_[1] challenge_[1] chapter_[1] community_[1] contact_[1] contribute_[1]
culture_[1] edit_[1] establish_[1] file_[1] finance_[1] illustrate_[1] image_[1]
indicate_[1] individual_[1] legal_[1] maintain_[1] percent_[1] policy_[1]

positive_[1] professional_[1] region_[1] register_[1] remove_[1] research_[1]
select_[1] specific_[1] tradition_[1] version_[1]

BNC-COCA-3,000 types: [fams 31 : types 31 : tokens 31]

advocate_[1] analyse_[1] civil_[1] component_[1] comprehensive_[1]
confine_[1] consequent_[1] decade_[1] define_[1] derive_[1] document_[1]
dominate_[1] factor_[1] federal_[1] foundation_[1] incorporate_[1]
institute_[1] integrate_[1] media_[1] pose_[1] promote_[1] psychology_[1]
regulate_[1] resolve_[1] retain_[1] significant_[1] submit_[1] symbol_[1]
trend_[1] via_[1] visible_[1]

BNC-COCA-4,000 types: [fams 3 : types 3 : tokens 3]

behalf_[1] contrary_[1] terminate_[1]

BNC-COCA-5,000 types: [fams : types : tokens]

BNC-COCA-6,000 types: [fams : types : tokens]

BNC-COCA-7,000 types: [fams 1 : types 1 : tokens 1]

negate_[1]

7.7 Frequency levels of total text

Text coverage by frequency level and AWL, measured in percentage of tokens for all words in text.

Variables: Frequency levels in percentage of total text	(Proper nouns) + K1- K2	K3	K4- K9	AWL vocab. in text
Textbook: Access to English	high-	freq.	mid-freq	
Topic: English as a Universal Language				
Tailored text: Divided by a Common Language:	90.76	3.45 (94.2)	3.17 (97.4)	2.3
Tailored text: A Global Language	92.24	3.90 (96.1)	2.42 (98.6)	2.3
Authentic text: Renaming English <i>Brisbane Times</i>	88.86	7.53 (96.4)	2.67 (99.1)	4.3
Topic: Indigenous peoples				

Tailored text: Native Americans: Original Inhabitants	90.47	5.79 (96.3)	2.19 (98.5)	3.6
Tailored text: Aboriginal Australians	89.97	4.69 (94.7)	3.29 (98)	4.1
Tailored text: Stolen Children	92.93	4.38 (97.3)	2.35 (99.7)	3.4
Authentic text: Native Americans In Business	90.35	5.63 (96)	1.76 (97.7)	8.4
Textbook: Stunt				
Topic: English as a Universal Language				
Tailored text: British vs. American English	91.23	2.51 (93.7)	4.48 (98.2)	2.4
Tailored text: English as a World Language	94.17	3.18 (97.4)	1.23 (98.6)	3.9
Authentic text: There Is an Epidemic UNESCO	89.89	5.97 (95.9)	3.09 (99)	8.2
Topic: Indigenous peoples				
Tailored text: Native Americans	90.78	4.78 (95.6)	2.77 (98.3)	3.6
Tailored text: Australia: The Birth of a Nation	92.80	4.72 (97.52)	2.24 (99.8)	2.6
Tailored text: Stolen Generation	93.06	4.56 (97.62)	2.18 (99.8)	5.6
Authentic text: Effects of Removal National Humanities Center	87.75	7.55 (95.3)	3.47 (98.8)	7.4
Textbook: Targets				
Topic: English as a Universal Language				
Tailored text: The Flavours of English	90.57	3.71 (94.3)	2.65 (96.9)	3.4
Tailored text: The Power of English Part 1	90.26	7.15 (97.4)	1.12 (98.5)	6.5
Tailored text: The Power of English Part 2	90.51	6.11 (96.6)	2.20 (98.8)	5.2
Authentic text: English and the Future BBC	92.52	5.58 (98.1)	1.44 (99.5)	7.3
Topic: Indigenous peoples				
Tailored text: Native Americans: We Are Still Here	86.95	6.88 (93.8)	3.17 (97.5)	5.6
Tailored text: Australia – the Island Continent	85.33	5.91 (91.2)	3.98 (95.2)	6.8
Authentic text: “Indian” Mascots	85.21	7.46 (92.7)	4.22 (97.3)	8.1

7.8 Norwegian Social Science Data Services (NSD)

Norsk samfunnsvitenskapelig datatjeneste AS
NORWEGIAN SOCIAL SCIENCE DATA SERVICES



Harald Hårfages gate 29
N-5007 Bergen
Norway
Tel: +47 55 58 21 17
Fax: +47 55 58 96 50
nsd@red.uib.no
www.nsd.uib.no
Org nr: 985 321 884

RESULTAT AV MELDEPLIKTTEST: IKKE MELDEPLIKTIG

Du har oppgitt at hverken direkte eller indirekte personopplysninger skal registreres i forbindelse med prosjektet.

Når det ikke registreres personopplysninger, omfattes ikke prosjektet av meldeplikt, og du trenger ikke sende inn meldeskjema til oss.

Vi gjør oppmerksom på at dette er en veiledning basert på hvilke svar du selv har gitt i meldeplikttesten og ikke en formell vurdering.

Til info: For at prosjektet ikke skal være meldepliktig, forutsetter vi at alle opplysninger som registreres elektronisk i forbindelse med prosjektet er anonyme.

Med anonyme opplysninger forstås opplysninger som ikke på noe vis kan identifisere enkeltpersoner i et datamateriale, hverken:

- direkte via personentydige kjennetegn (som navn, personnummer, epostadresse e.l.)
- indirekte via kombinasjon av bakgrunnsvariabler (som bosted/institusjon, kjønn, alder osv.)
- via kode og koblingsnøkkel som viser til personopplysninger (f.eks. en navneliste)
- eller via gjenkjennelige ansikter e.l. på bilde eller videoopptak.

Vi forutsetter videre at navn/samtykkedeklæringer ikke knyttes til sensitive opplysninger.

Med vennlig hilsen,

NSD Personvern