Statistical Learning and Inconsistent Language Input: An Experiment in Grammar Learning.

Andrea Eide and Elisabeth Skeistrand

Master's thesis

Master's Programme in Health Sciences,

Speech-Language Pathology

Department of Biological and Medical Psychology

The Faculty of Psychology

University of Bergen

Spring 2019

Acknowledgements

The work with this thesis has been challenging, but also exiting. Conducting this experiment was much more extensive than we could have imagined, and we have learned a lot. At times the process has been frustrating, and we had to rethink and start over again several times. In some periods, we have been in the laboratory from the morning to the late evening. This experience included a lot of fun as well, and we managed to reach the finish line. We thank each other for the countless hours spent on this master thesis, and for both of us being positive and supportive of each other during the whole process.

A big thank you to the participants in our pilot study for your time and excellent feedback, which helped us improve the experimental design. We also want to thank everyone who volunteered and participated in the final experiment. The experiment required a lot of participants, and the recruitment process was challenging at times. We are so grateful for everyone who devoted their time to participate. There would be no experiment without you.

We would like to thank our supervisors, Dr. Arve Asbjørnsen and Phd. Candidate Sunniva Sørhus Eidsvåg at the University of Bergen, for all their valuable feedback and advice during the work with our master thesis. We are so grateful for your engagement, many helpful discussions, and for all the hours you have spent helping us throughout the process.

At last, we thank our families and our partners for their patience, support and for cheering us all the way to the finish line.

The experiment was funded by the Department of Biological and Medical Psychology at the University of Bergen.

Table of contents

Abstract

Our master thesis is based on an experiment, conducted to investigate the resilience of the statistical learning mechanisms. We familiarized adult participants to noun gender subcategories in an unfamiliar language (Russian), with different amounts and types of inconsistencies in the input. A theoretical background of statistical learning research is presented in the first part of the thesis. This includes theory on language acquisition, grammar learning and inconsistent input, as well as clinical implications. In the methodology section, we present our research design, participants, materials, procedures, methodological challenges and analyses. Moreover, we discuss the ethical aspects of the experiment, as well as considerations regarding reliability and validity. Fifty-six participants (28 male and 28 female) were allocated in two experimental groups (High and Low Inconsistency). The familiarization phases differed for the two groups, as the stimuli material for each group contained different percentages of inconsistent items (25% and 50%). In the following test phases, the participants were tested on novel lexical items, to assess generalization. Familiarization and testing were completed three times in total, in three Stimuli Conditions. The conditions contained different types of stimuli (Grammatical items and Ungrammatical items). Participants in both experimental groups showed a statistically significant learning effect, suggesting abilities to learn gender subcategories even with high levels of inconsistent input. Additionally, the learning effect was significant in all Stimuli Conditions, not differentiating between familiarization of grammatically correct or incorrect input.

Key words: Language acquisition, Statistical learning, Inconsistency

Sammendrag

Denne masteroppgaven bygger på et eksperiment som ble utført for å undersøke robustheten ved statistisk læring. Voksne deltakere ble familiarisert med maskuline og feminine substantiver med doble kjønnsmarkeringer fra et ukjent språk (russisk). Materialet inneholdt ulike mengder og typer av ugrammatiske elementer. I oppgavens teoridel vil vi gi en oversikt over den teoretiske bakgrunnen for forskning innen statistisk læring. Dette inkluderer teori om språkinnlæring, læring av grammatikk og inkonsistent input. I metodedelen presenterer vi vårt forskningsdesign, deltakere, materiale, prosedyrer, metodologiske utfordringer og analyser. Avslutningsvis blir etiske aspekt og hensyn til validitet og reliabilitet diskutert. Femtiseks voksne (28 menn og 28 kvinner) ble fordelt i to eksperimentelle grupper (Høy og Lav Inkonsekvens). Familiariseringsfasene for de to gruppene var ulike, ved at stimuli materialet for hver gruppe inneholdt ulike prosentdeler av ugrammatiske elementer (25% og 50%). I den følgende testfasen ble deltagerne testet i nye leksikalske enheter for å undersøke deres evne til å generalisere. Familiarisering og testing ble gjennomført tre ganger, i tre ulike stimuli betingelser. Hver betingelse inneholdt ulike typer stimuli (ord som var grammatisk korrekt, ord som hadde feil kjønnsmarkering og ord som ikke hadde kjønnsmarkering). Deltagerne i begge grupper (Høy og Lav Inkonsekvens) viste signifikant læringseffekt, noe som indikerer evnen til å lære grammatiske subkategorier, også ved en høy andel av feil input. Læringseffekten var signifikant i alle betingelser, noe som indikerer at det ikke var forskjeller mellom familiarisering av grammatisk og ugrammatisk input.

Nøkkelord: språkinnlæring, statistisk læring, inkonsistent input

Statistical Learning and Inconsistent Language Input: An Experiment in Grammar Learning.

Our experiment was designed to explore the mechanisms of language learning, based on the research traditions of statistical and implicit learning. The purpose was to investigate adult learners' tolerance of inconsistencies in their language input.

In this thesis, we will present an overview of the theoretical and historical background of literature on language acquisition, and further describe the perspectives of the implicit and statistical learning traditions. Moreover, relevant literature from these traditions will be presented, such as grammar learning and the role of inconsistent language input, as well as clinical implications. Finally, we introduce the methodological background, including a discussion of methodological challenges, ethical aspects, reliability and validity.

Theory

Language Acquisition

Language acquisition is a highly complex process, because of the multiple levels of intricate structures in natural languages. These levels of structure include the complexity of distributions within phonemes, words, phrases and utterances (Romberg & Saffran, 2013). The underlying mechanisms of language acquisition has been debated and explained by different theoretical perspectives. A fundamental difference between some of the theories, is whether learning plays an important role, or if acquisition of language is an innate ability (Saffran, 2003).

The linguist Noam Chomsky is considered the front figure of the Nativist tradition. This theoretical perspective is based on the belief that language acquisition is innate, meaning that humans are born with a great deal of pre-existing linguistic knowledge which does not require learning (Chomsky, 1965). According to Chomsky (1965), this innate ability is linked to a specific module for language in the brain, and language acquisition is explained as a maturation of this module. The human ability to learn languages is considered innate, which explains why all languages in the word share some commonalities. This constitutes a common language system known as “Universal Grammar” (Chomsky, 1965).

Chomsky claimed that the ability to learn language is distinct from other cognitive abilities (Chomsky, 1965). This view has later been challenged by the learning-oriented tradition, which claims that humans possess powerful learning mechanisms, allowing us to detect statistics and patterns in our environment without any instructions. Importantly, the learning-oriented tradition argues that these learning mechanisms are not restricted to language acquisition alone, suggesting them to apply for all processing of information in our environment. Additionally, similarities across different languages can be explained as a result of constraints on learning, as these general learning mechanisms shape our languages (Saffran, 2003). Two different research traditions have both provided evidence for these domain-general learning mechanisms: the implicit learning tradition and the statistical learning tradition.

The Implicit Learning and Statistical Learning Traditions

Implicit Learning. The implicit learning tradition began with the early work of Reber (1967), who presented one of the first influential experiments using an artificial grammar learning task. Letter strings were generated by using a finite state grammar, consisting of five letters (*P*, *S*, *T*, *V* and *X*). Adult participants were exposed to the input visually, and asked to memorize the sentences they had looked at. After the familiarization, the participants were informed that the letter strings had followed a set of grammatical rules. Then they were presented with new grammatical and ungrammatical letter strings, and asked to judge whether these strings were correct or not. The results from this influential study showed that the participants were able to distinguish between grammatical and ungrammatical strings, indicating that they had acquired the unfamiliar artificial grammar rules (Reber, 1967).

However, they were not able to describe the underlying rules, and most of them did not feel like they had learned anything during the experiment. Based on this, Reber (1967) concluded that they lacked conscious knowledge, and introduced the term implicit learning to describe this unconscious learning mechanism.

This study was based on the Artificial Grammar Learning-paradigm (AGL). Other paradigms within the implicit learning tradition are Sequence Learning (SL) and Dynamic System Control (DSC). In all these paradigms, the participants are typically exposed to a complex and rule-based environment. These studies also include measuring the participants' ability to verbally express the knowledge they have (or have not) acquired, as well as measures of consciousness (Cleeremans, Destrebecqz, & Boyer, 1998).

Originating from cognitive psychology, the distinctions between implicit and explicit learning mechanisms are largely debated. The role of consciousness in learning have been a central matter in these discussions (Ellis, 2009). Implicit learning has been suggested to occur without awareness or intention, while explicit learning is fully conscious and hypothesis-driven (Cleeremans et al., 1998). The mechanism of implicit learning, in the sense of proceeding without awareness or consciousness, is highly debated. Shanks (2003, p. 36) argues that there is not enough evidence separating implicit from explicit learning, and proposes the possibility of a single underlying learning mechanism. Cleeremans et al. (1998) argue that learning always will be accompanied with a certain degree of awareness, presenting the lack of a proper operational definition of awareness as the main conceptual problem in implicit learning research.

Regarding second language acquisition, Schmidt (1990) defined consciousness in three terms, distinguishing consciousness as awareness, as intention and as knowledge. Consciousness as awareness would thus make a distinction between implicit and explicit learning. Continuing this, two levels of awareness were proposed: awareness as simple noticing or as metalinguistic awareness (Schmidt, 2001). Language learning will usually involve awareness in simple noticing of surface elements. In this sense, implicit learning can be defined as learning without metalinguistic awareness, which excludes the metalinguistic levels of analysing underlying rules (Ellis, 2009).

In the present experiment, we decided to briefly investigate the awareness of the participants through subjective measures in a self-rating questionnaire. The main purpose was to assess potential use of explicit learning strategies, like fully conscious hypothesis-testing. Nevertheless, results from self-raring questionnaires have their limitations, and the self-reported degree of awareness was accordingly interpreted in a speculative manner.

Statistical Learning. The term “statistical learning” was introduced by Saffran, Aslin and Newport in 1996, and is described as the mechanism that enables both adults, children and infant to extract patterns of statistics from our environment (Aslin & Newport, 2012). In terms of language acquisition, Saffran (2003) emphasizes how learners can extract phonological, grammatical and syntactic structures from linguistic input by using statistical properties. Statistical learning is considered as a domain-general mechanism, where learning has been demonstrated with tactile stimuli (Conway & Christiansen, 2005) and visual stimuli (Fiser & Aslin, 2002; Fiser, Aslin, & Lindsay, 2005; Kirkham, Slemmer, & Johnson, 2002). In addition, these mechanisms do not seem to be unique to the human race, in which studies with both rats (Toro, Trobalon, Sebastián-Gallés, & Mackintosh, 2005) and cotton-top tamarin monkeys (Saffran et al., 2008) have provided evidence of statistical learning.

Studies within the tradition of statistical learning are characterized by thorough manipulation of the statistical information in the input, mainly by using artificial languages (Rebuschat & Williams, 2012). Typically, the learners are presented with input for several minutes and subsequently tested on what they have learned, without instructions or feedback during the experimental tasks (Plante & Gómez, 2018). It has been claimed that statistical learning and rule learning are two separate mechanisms, due to the former involving learning the elements presented during exposure, and the latter entailing generalization to novel instances (Aslin & Newport, 2012). Aslin and Newport (2012, p. 170) present a unifying perspective that argues for “a single statistical-learning mechanism that accounts for both the learning of input stimuli and the generalization of learned patterns to novel instances”. We based our experiment on this approach, concerning statistical learning and rule learning as different outcomes of the same mechanism.

Research within the statistical language learning paradigm have traditionally focused on infant language acquisition, but several studies with adult participants have also been conducted throughout the last two decades (Rebuschat & Williams, 2012). In most of the infant studies, the Head-turn Preference Procedure has been used during testing (for explanation, see Kemler Nelson et al., 1995).

Perruchet and Pacton (2006) argues that the implicit and statistical learning traditions in fact are studying the same, and that the divergence between these approaches will lead to a major theoretical challenge for future studies. Similar to the implicit learning tradition, the statistical learning approach focuses on the human acquisition of information from the surrounding environments, traditionally by using artificial grammars (Hamrick & Rebuschat, 2012). However, the two approaches differ from each other in some manners. Implicit learning research have mainly focused on the formation of chunks, while the statistical learning tradition has focused on statistical computations. Additionally, statistical learning research have primarily investigated the acquisition of linguistic input, while the implicit learning tradition explores learning in several domains (Hamrick & Rebuschat, 2012; Perruchet & Pacton, 2006). Despite some differences, Perruchet and Pacton (2006) suggest that the two research traditions are ultimately exploring the same domain-general incidental learning process. Our experiment is based on earlier findings from both traditions, yet we have chosen to apply "statistical learning" as a general term, in line with similar studies (Eidsvåg, Austad, Plante, & Asbjørnsen, 2015; Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b).

Studies on Statistical Language Learning

One of the first challenges the infant language learner must face, is how sequences of sounds or syllables in speech streams represent words. To do this, they must detect transitional probabilities (also known as conditional probabilities), which can be explained as the likelihood of one linguistic element predicting the occurrence of another element (Plante & Gómez, 2018).

In their early influential study, Saffran et al. (1996a) investigated whether 8-month-olds were able to use statistical information to identify word boundaries in fluent speech. During familiarization, the infants were exposed to a continuous speech stream from an artificial language for two minutes. This speech stream consisted of four three-syllable nonsense words (e.g., *bidakupadotigolabubidaku...),* and was repeated in random order. The transitional probabilities between pairs of syllables were the only cues to word boundaries in the auditory input, and was higher within words (1.0), than between words (.33). The test phase was carried out by using the Head-turn Preference Procedure (Saffran et al., 1996a). Saffran et al. (1996a), and found that infants were able to discriminate between both word and non-word stimuli, in which they had longer listening time for non-words during testing. In a second experiment, infants were also able to distinguish words from part-words, which required them to exploit an even more difficult statistical computation. These results provided evidence that infants are sensitive to the transitional probabilities of adjacent syllables, and that they can organize presented auditory input based on differences in the transitional probabilities. Additionally, Saffran et al. (1996b) found similar evidence of learning in adults, including that the presence of prosodic cues seems to enhance adults' performance.

Since then, researchers have continued to investigate statistical learning, and in many later studies the complexity of the artificial languages have been increased to strengthen their ecological validity (Arciuli & Torkildsen, 2012; Hay & Lany, 2012). For example, Gómez and Gerken (1999) found evidence of infants being able to segment words with a lower range of transitional probabilities (e.g., mean transitional probability for grammatical test items was 0.42 in Experiment 1 and 2, 0.395 and 0.6 in Experiment 3), as in natural languages. Importantly, infants (8-month-olds) ability to segment words from fluent speech has also been documented in studies using natural language stimuli (Italian) (Pelucchi, Hay, & Saffran, 2009). Moreover, research findings have demonstrated that previous exposure to word forms, in a typical statistical word-segmentation task, facilitates subsequent learning of word meaning in both infants (Graf Estes, Evans, Alibali, & Saffran, 2007) and adults (Mirman, Magnuson, Estes, & Dixon, 2008).

Grammatical Dependencies. Grammatical dependencies between elements that co-occur is called adjacent elements. Sentences in natural languages contain such adjacent dependencies, in addition to non-adjacent dependencies. These are dependencies between elements that do not co-occur (Wang, Mintz, Greene, & Benjamin, 2018). For example, the dependency between *the boy + cries* are adjacent, but becomes non-adjacent if the sentence is changed to *the boy* + always + *cries.* Research regarding sensitivity to non-adjacent dependencies has provided varied results and suggest that learning of these kinds of structures is more difficult than learning of adjacent dependencies (Newport & Aslin, 2004). This is because elements must be remembered for a longer time in order to be linked to elements occurring later. Additionally, there will be one (or more) elements in between the dependent elements, potentially disturbing the learners tracking of the relevant dependencies (Hay & Lany, 2012). However, the presence of specific conditions like high variability in the adjacent dependencies (e.g., Gómez, 2002; Gómez & Maye, 2005) and the appearance of silences or edges (i.e., the dependent elements are at the edges) between the non-adjacent dependencies will enhance the learning (e.g., Newport & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002; Wang et al., 2018).

An issue of particular interest in the later years, has been the effect of input variability. Especially regarding the demonstration of enhanced learning of non-adjacent dependencies, after the influential study by Gómez (2002). In her study, both infants and adults increased their ability to compute statistical relations between non-adjacent dependencies (e.g., *aXd*, *bXe*), as the variability in the middle element (*X*) increased. When the variability of the *X*-element increased, the transitional probability between the adjacent elements (e.g., *a*+*X* in an *aXd* sentence) decreased. Gómez (2002) suggested that the high input variability was making the non-adjacent dependencies more salient. The same effect has been replicated in several later studies, strengthening the hypothesis that high variability facilitates learning (e.g., Aguilar, Plante, & Sandoval, 2018; Eidsvåg et al., 2015; Gómez & Maye, 2005; Grunow, Spaulding, Gómez, & Plante, 2006; Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013).

The Role of Statistical Cues. Learning grammatical structures require learning of different grammatical categories, such as nouns, verbs and determiners, and how they can be combined. Distributional and phonological cues are some common statistical cues that can promote the learners' ability to form grammatical categories (Sandoval, Gonzales, & Gómez, 2012). Distributional cues can be explained as the syntactic context in which a particular word is likely to occur, while phonological cues are the sound properties of a word (Mintz, 2002). It has been discussed whether language learners are able to use distributional cues alone to detect grammatical categories, or whether the presence of multiple correlated cues are required (Hay & Lany, 2012). Although previous research findings have demonstrated that adults are able to learn grammatical-like categories in artificial grammar learning tasks from distributional cues alone (e.g., Reeder, Newport, & Aslin, 2013, 2017), Sandoval et al. (2012)

emphasizes that word class acquisition cannot solely be based on the presence of one particular kind of statistical cue. This is because languages contain multiple correlated cues to word classes.

In an early influential study, Braine (1987) explored adults learning of gender subcategories when multiple correlated cues were present. In this study, adults were exposed to an artificial grammar, consisting of the four elements *M*, *N*, *P* and *Q*. These elements were put together in two sets of co-occurrences (distributional information), *MN* and *PQ*. Previous findings had suggested that adults were able to learn that *M* and *P* appeared first in phrases, but they had not detected the co-occurrence between *M* and *N*, or *P* and *Q*. Braine (1987) hypothesized that this could be related to the lack of multiple cues, and he therefore added a semantic cue which provided referential information (gender) to half of the *M*- and *P*-words. The *M*-words were paired with pictures of men, while the *P*-words were paired with pictures of women. The other half of the *M*- and *P*-words were paired with pictures of non-living objects, unrelated to a semantic category. The *N*- and *Q*-words indicated the number of nouns in the picture, e.g., “one”, “two”. For the control group, gender was not correlated with the *M*- and *P*-words. Adults in the experimental group correctly judged the grammaticality in both familiar and novel phrases, even when the phrases were paired with pictures of non-living objects. After exposure to both distributional information and semantic cues, the participants in the experimental group formed categories of the four presented elements. They also demonstrated the ability to generalize, by some of the pairings in the testing being novel. Braine (1987) explained this learning process as step-based. Step 1 learning is the ability to discover the co-occurrence between *MP* and *NQ* with the presence of distinguishing cues. Step 2 learning is when the learners can generalize to novel instances, without the appearance of distinguishing cues (Braine, 1987). Later research findings from similar studies have revealed that adult learners are able acquire and generalize gender subcategories in artificial languages based on the presence of multiple phonological cues as well (e.g., Brooks, Braine, Catalano, Brody, & Sudhalter, 1993; Frigo & McDonald, 1998).

Gómez and Lakusta (2004) investigated such Step 1 learning (Braine, 1987) in their study exploring whether infants could learn dependencies between adjacent word categories. In the first of two experiments, 12-month-old infants were familiarized with strings consisting of *aX* and *bY* pairing, or *aY* and *bX* pairing. The *a*-, *b*-, *X*- and *Y*-elements represented different word categories and were combined to form phrases. The *X*-words were monosyllabic, and the *Y*-words were disyllabic, serving as phonological cues. After 3 minutes of exposure, the infants were tested on new phrases from their training language, as well as phrases from the other language. Infants were able to discriminate between legal and illegal pairings while introduced with in novel *X*- and *Y*-words (Gómez & Lakusta, 2004). Their findings indicate that 12-month-old infants are able to group elements from artificial languages into categories (*X*- and *Y*-elements). They also suggested that infants can form associations between other elements (*a*- and *b*-elements) and these categories (Gómez & Lakusta, 2004).

In the present experiment we investigate adults' ability to learn gender subcategories in an unfamiliar language (Russian). We included items with double gender markings, serving as multiple morphophonological cues. Previous studies have investigated both infants' and adults' ability to learn Russian grammars within similar stimuli items. For example, Gerken, Wilson, and Lewis (2005) familiarized 17-month-old English-learning infants with a Russian gender paradigm, consisting of both single and double morphophonological cues. The stimuli material used during training consisted of six masculine and six feminine Russian lexical roots with two different gender markings for each. Some of the lexical roots were inflected with a single gender marking, while other had double gender markings. After only two minutes of exposure, infants were able to distinguish novel grammatical Russian words from ungrammatical words (unmarked words), indicating Step 2 learning (Braine, 1987). However, the results from this study revealed that the infants were only able to do so when a subset of the familiarization items consisted of double gender markings, serving as additional morphophonological cues (Gerken et al., 2005).

Results from similar studies have demonstrated that adult learners can learn Russian gender subcategories, based on the presence of distributional information and morphophonological cues (e.g., Richardson, Harris, Plante, & Gerken, 2006; Wilson, Gerken, & Nicol, 2002). Eidsvåg et al. (2015) extended these previous findings by investigating whether increased variability would enhance the learning, as indicated by previous research (e.g., Gómez, 2002). As in the previous studies, adults in this study were able to distinguish novel grammatical Russian words from ungrammatical words, and the presence of variability in the input enhanced the participants learning outcome. As suggested by previous literature (e.g., Gerken et al., 2005), the double-marked words were easier to learn than the single-marked words, due to the presence of multiple subcategory cues (Eidsvåg et al., 2015).

Inconsistent Input in Natural Language Environments

Experiments within the statistical learning paradigm are, as mentioned, typically exposing learners to consistent statistical cues from artificial language structures. Learners must detect the statistical properties to learn the underlying grammatical structure, and the learning outcome can be measured by the participants' ability to generalize their knowledge to novel instances. In a natural language environment, the learner is exposed to language input adhering to the consistent grammatical rules, as well as ungrammatical input, which is violating these rules. In the present study, we have chosen to apply the term "inconsistency" when referring to elements that do not adhere to the consistent grammatical rules of a language.

Inconsistencies in the input can be both sporadic and systematic (Gómez & Lakusta, 2004). In natural language acquisition, infants get their language input from multiple different talkers with varied pronunciation, vocabulary and grammar. This is due to different factors like gender, dialects and proficiency (Gonzales, Gerken, & Gómez, 2018). Inconsistent input in the forms of erroneous articulation and inflection, are frequently heard in children (particularly with language impairments) and second language learners (Bulgarelli, Lebkuecher, & Weiss, 2018; Montgomery & Leonard, 1998).

Alt (2018) describes statistical learning as a robust mechanism. However, most of the previous research conducted to investigate statistical learning have used language input adhering to a consistent set of grammatical rules during familiarization. A question of particular interest is how robust these mechanisms are when the learners are presented with a subset of items that violates the grammatical rules. How much inconsistency can the learners tolerate before the learning outcome is reduced?

Previous studies have explored this robustness of statistical learning by exposing 12-month-old infants to two different artificial languages, containing different amounts of inconsistent items (Gómez & Lakusta, 2004; Gonzales, Gerken, & Gómez, 2015; Gonzales et al., 2018). Inconsistent items have typically been created by inserting items from the other artificial language, which violated the abstract grammar rules from the infants' training language. The results have revealed that infants are able to learn the abstract grammar rules from their training language when presented 17% inconsistency (Gómez & Lakusta, 2004), or even higher percentages (25% and 38%) when the different input streams were presented by different voices (Gonzales et al., 2018), or separately (Gonzales et al., 2015). Studies using adult participants have found that adult learners tend to acquire and reproduce their input as presented to them, including the inconsistent items (Hudson Kam & Newport, 2005, 2009). However, when the inconsistent items were presented with a high degree of variability, the adults' learning of the consistent grammar rules increased as well (Hudson Kam & Newport, 2009).

Traditionally, research regarding statistical learning have used artificial language stimuli. Artificial languages permit greater control of all variables, but a disadvantage is the difficulty of knowing whether the findings can be generalized to a natural language environment. In later years, there has been done several studies on statistical learning with natural language stimuli as well (e.g., Eidsvåg et al., 2015; Gerken et al., 2005; Kittleson, Aguilar, Tokerud, Plante, & Asbjørnsen, 2010; Pelucchi et al., 2009)

By using inconsistent structures which are frequently found in the natural language environment, the ecological validity is enhanced. Several studies on grammar learning have demonstrated the use of common inflectional errors as inconsistent input, such as omission errors (Hudson Kam & Newport, 2005) and substitution errors (Eidsvåg et al., 2015; Hudson Kam & Newport, 2009). Omission errors are defined as the missing of elements which should be present (Kusumawardhani, 2017) such as absent grammatical markers in lexical roots that require inflection. Substitution error is the term for an incorrect replacement of words or phonemes (Santos, 2012), such as inflecting a lexical root with an incorrect grammatical marker.

In the present experiment, we chose to pursue the use of Russian language stimuli, consistent with Gerken et al. (2005) and Eidsvåg et al. (2015). The purpose was to investigate whether different amounts of inconsistency in the input would affect the learning outcome, as well as exploring the effects of different types of inconsistencies. We chose to include substitution and omission errors as inconsistent items.

Inconsistent Input and Bilingualism. The previous presented studies regarding inconsistency, have focused on monolingual infants and adults. Bilingual infants and children are exposed to multiple languages at the same time. They seem to acquire their languages in a similar manner as their monolingual peers, despite an additional amount of inconsistent language input (Bulgarelli et al., 2018). A recent study by Antovich and Graf Estes (2018) compared bilingual and monolingual infants’ ability to track regularities from dual interleaved speech streams. They found that only the bilingual infants were able to do so, if an indexical cue was present to indicate the presence of multiple streams (Antovich & Graf Estes, 2018). Moreover, de Bree, Verhagen, Kerkhoff, Doedens, and Unsworth (2017) conducted the first study, of our knowledge, comparing monolingual and bilingual toddlers’ ability to learn non-adjacent dependencies. Moreover, they explored whether the learning outcome would be affected when a subset of the input material contained inconsistent items. Their results showed that only the bilingual toddlers demonstrated learning of the non-adjacent dependencies when 14% of the strings were inconsistent items (de Bree et al., 2017). Findings from Gonzales et al. (2018) also indicated a bilingual advantage when comparing the results by infants with no more than 10 hours of exposure to a second language, and infants with greater exposure.

Altogether, the results from these previous studies indicate that bilingual infants and toddlers might benefit from their experience with more complex and varied input, when tracking statistical information. Based on these tendencies, we chose to collect information about the participants general language knowledge. This was assessed in a self-rating questionnaire, in order to explore whether the participants language background might affect their ability to learn from inconsistent input.

Clinical Implications

The literature mentioned above, emphasize the importance of statistical learning in language acquisition and processing. Several studies have suggested that children with learning disabilities and Developmental Language Disorders (DLD), are not able to detect statistical structures as easily as typically developing children (Evans, Saffran, & Robe-

Torres, 2009; Haebig, Saffran, & Ellis Weismer, 2017). Similar findings have been shown in adults, where participants with a history of language-based learning disabilities performed below adults with normal language, when extracting statistical properties of an underlying grammatical system (Plante, Gomez, & Gerken, 2002; Richardson et al., 2006). This is a major concern for speech-language pathologies (SLPs), when treating individuals with language-related difficulties.

Principles based on theoretical knowledge from the statistical learning research are useful in designing intervention strategies for children with DLD, as demonstrated by Alt, Meyers, and Ancharski (2012). An important aspect is the principle of input variability, which enhances the learning of inherent patterns in the language, rather than exemplar-based learning. Correspondingly, Torkildsen et al. (2013) found that adults with language-based learning disabilities, placed in a high variability group, were able to demonstrate generalization of the underlying grammar. This is supported by a body of research on sample populations of individuals with normal language (Eidsvåg et al., 2015; Gómez, 2002). Alt et al. (2012) suggest that impaired language learners do not use different learning methods than learners with normal language, yet they are less efficient in their learning. Based on research on unimpaired learners, incorporation of high variability in treatment interventions for language-related difficulties can thus be recommended.

More recent studies have suggested similar benefits in the treatment of individuals with language-based learning disabilities as well. The advantages of statistical learning principles include quick and efficient learning, generalization of learning, and focus on input, which reduce the behavioural demands on the learner (Alt, 2018). Elements from statistical learning ought to be incorporated into language therapy interventions, based on applicable evidence for the work of a SLP. Recent evidence has shown that children with DLD benefit from grammatical morpheme treatment using these principles (Owen Van Horne, Curran, Larson, & Fey, 2018). Additionally, children and college students with DLD were shown to form grammatical categories, engaging in these strategies (Hall, Owen Van Horne, McGregor, & Farmer, 2018). Children with autism spectrum disorder learned verb meanings in clinical use of the same type of principles as well (Horvath, McDermott, Reilly, & Arunachalam, 2018).

Lian (2017) explored some important aspects regarding the domain-general mechanism of statistical learning, in respect to diagnosis and treatment of language impairments. He suggested the possibility of non-linguistic training tasks as potential remedial treatment. This recommendation was based on research on domain-general learning abilities, showing that the enhancement of these general abilities in hard-of-hearing children, additionally served to improve functions of language (e.g., Conway, Gremp, Walk, Bauernschmidt, & Pisoni, 2012). Finally, Lian (2017) emphasized the need for additional research to make any conclusions. Further studies on statistical learning as a domain-general learning mechanism may contribute to future treatment of language impairments.

Hsu and Bishop (2010) highlight the need for further empirical examination as well, especially in the context of language impairment. Future development of treatment interventions will be enhanced by increasing the empirical basis regarding the settings that facilitates statistical leaning.

In the present study, we explored how much inconsistent language input learners can tolerate. Erroneous input from the language environment (like substitution and omission errors), is a highly relevant concern in the work of SLPs. As mentioned, these types of grammatical inconsistencies are commonly heard in children with language impairments. Experiments like ours can contribute with further understanding, in regard to whether high levels of inconsistent input might disrupt the language learning process. By conducting studies with the purpose of investigating the robustness of the statistical learning mechanisms, we wish to supply the body of research on the broad topic of language acquisition. The basis of evidence on normal learners will further be beneficial for continuing research on language acquisition in individuals with DLD. There is enough evidence on statistical learning principles to start incorporating these into practice, as suggested by Alt (2018), yet the need for more research is evident.

Research Questions and Hypotheses

In the present experiment, the main objective was how much inconsistent input adult learners are able to tolerate. We also wanted to explore whether different types of inconsistencies might affect the learning outcome. Adult participants were allocated in two experimental groups and familiarized with Russian nouns in three Stimuli Conditions. In one of the conditions, the familiarization consisted of grammatically correct nouns. In the two other conditions, the familiarization stimuli included inconsistent items (substitution or omission errors). The percentage of inconsistent items differed between the groups, with respectively 25% (i.e., Low Inconsistency group) or 50% (i.e., High Inconsistency group) inconsistency.

The following hypotheses were formulated based on the research questions:

- Lower percentage of inconsistent input (25%), will result in a significantly greater learning outcome compared to a higher percentage (50%).
- The performance in both groups will be greater after exposure to the Grammatical Condition (only grammatically correct items), due to the absence of inconsistent items in the input.

Methodology

Research Design

We have chosen a quantitative research design to answer our research questions. Quantitative approaches are traditionally used to find in-depth knowledge, causal relationships and to test if the stated hypotheses can be generalized (Drageset & Ellingsen, 2009). In quantitative research, numeric information is collected, which allows us to analyse the gathered data statistically (Polit & Beck, 2017). The present design is based on the traditional positivist scientific method, directed at investigating underlying causes of a phenomenon by gathering empirical evidence through objective and disciplined procedures (Polit & Beck, 2017).

Quantitative research traditions distinguish between experimental and nonexperimental designs. An experimental design is characterized by the researchers being active agents in the experiment, meanwhile in nonexperimental designs the researchers contributes as passive observers (Polit & Beck, 2017). Experimental methods measure responses by permitting control of some variables and direct manipulation of the variable of interest (Cozby & Bates, 2015). To answer our research question, an experimental design was considered appropriate.

Recall that in the present study, participants in two experimental groups (High Inconsistency and Low Inconsistency) were exposed to different amounts of inconsistencies in the input. The stimuli material was presented to the participants in three Stimuli Conditions (referred to as Grammatical Condition, Substitution Condition and Omission Condition), each consisting of a familiarization phase immediately followed by a test phase. In two of the familiarization phases (Substitution Condition and Omission Condition), the stimuli set included inconsistent items: 50 % inconsistent items for the High Inconsistency group and 25% inconsistent items for the Low Inconsistency group. The remaining familiarization phase (Grammatical Condition) included no grammatical errors, serving as a control.

The independent variables in this experiment were Experimental Group (High Inconsistency and Low Inconsistency) and Stimuli Condition. These were manipulated in order to consider the outcome in the dependent variable of Acceptance Rate. Furthermore, the independent variables of previous language skills and degree of awareness was additionally measured, based on data collected in a self-rating questionnaire after the main experiment.

Participants

As mentioned, a body of research have used children and infants as participants to investigate the mechanisms of statistical learning (e.g., Gerken et al., 2005; Gómez, 2002; Gómez & Lakusta, 2004; Pelucchi et al., 2009; Saffran et al., 1996a). In our experiment, we wanted to explore whether these tendencies also apply for adult participants.

A total of 56 adults (ranging from 18 to 56 years old) were included in the experiment. Twenty-eight males (mean age of 25.0, *SD* 6.7) and twenty-eight females (mean age 23.7, *SD* 5.1) were equally arranged in each of the two groups (High Inconsistency and Low Inconsistency). Participants were allocated in the groups by pseudo-randomization (same number of males and females in each group), in an attempt to control for participant sex.

Norwegian as native language and age above 18 years old were part of the inclusion criteria. Previous research on implicit and statistical learning, have found distinctions regarding the learning effects in individuals with learning disabilities, developmental language disorders and autism spectrum disorder (Evans et al., 2009; Haebig et al., 2017). Considering this, we decided to include a sample of a general adult population, making the exclusion criteria to be known hearing loss, language deficits, and developmental or acquired neurological disorders. Prior knowledge of Russian, or other related Slavic languages, were additional criteria of exclusion.

For recruitment, purposive sampling was conducted. Purposive sampling is a form of nonprobability sampling, where participants are included based on predetermined inclusion criteria (Cozby & Bates, 2015). This enabled a cheap, efficient and convenient recruitment process, but has its limitations regarding the generalizability of the study. During the autumn of 2018 the recruitment process was conducted. Posters and flyers were put up on message boards, placed in cafeterias and in the libraries at the various faculties at the University of Bergen. Additionally, we posted information about the project in various social communities on Facebook at several occasions. The inclusion and exclusion criteria were specified in all instances. All of the volunteers were contacted and invited to participate in the experiment, until we reached the desired number of participants.

The participants received a compensation of NOK 150 (approximately USD 20) to cover any potential expenses associated with the participation. After completing the experiment, they received the money either by bank transfer, cash or by the application Vipps (Norwegian digital payment service for smartphones). All subjects signed a form after receiving the compensation by their preferable payment method. There was no withdrawal of any participants.

Materials and Procedures

The stimuli material in the present experiment is a modified version of the Russian stimuli previously used by Gerken et al. (2005), Richardson et al. (2006), Eidsvåg et al. (2015) and Sandoval, Patterson, Dai, Plante, and Vance (2017). We chose to use auditory stimuli items consistent with Eidsvåg et al. (2015) and Sandoval et al. (2017), only including the items with double gender markings. Double-marked items have been shown to facilitate a greater learning effect relative to single marked items, because of the presence of multiple morphological cues (Eidsvåg et al., 2015; Gerken et al., 2005; Richardson et al., 2006). The material is a highly simplified selection of a Russian inflectional paradigm, only consisting of feminine and masculine nouns. This simplification was necessary in adjustment to the framework of our master thesis, because of the complexity of the Russian word formation system. In Russian morphology, suffixation of nouns is a particularly complex process, with over a hundred different variations of suffixes and combinations (Wade & Gillespie, 2011).

Our stimuli material consisted of feminine and masculine Russian nouns with two potential double markings for each gender category. The masculine lexical roots had the gender markings *-tel+ya* and *-tel+yem*. For the feminine roots, the double-markings were *-k+oj* and *-k+u*. To demonstrate, the feminine lexical root *Blondin* could be inflected by both of the associated gender markings (e.g. *Blondin+k+oj* or *Blondin+k+u*). For the inconsistent stimuli items, some lexical roots were paired with an incorrect combination of gender markings (i.e., substitution errors: *-k+ya/yem*and *-tel+oj/u*), while some lexical roots did not hold any suffixes (i.e., omission errors: Ø).

The auditory stimuli material consisted of one audio file for each unique item. In the familiarization phases, the audio files were recorded by two different native Russian speakers (one male and one female). In the test phases, the stimuli items were presented by a novel Russian speaker (female). Prior to the present experiment, the audio files were edited by filtering out background noise to achieve the best sound quality possible. Some of the audio files were additionally edited, by removing the suffixes for use as omission items. Editing of the audio files was conducted in GoldWave Version 6.35 (GoldWave Inc, 2019).

All stimuli items were presented to the participants in random order, with a 300-580 milliseconds break between. The familiarization stimuli in all Stimuli Conditions consisted of the same 32 lexical roots. To ensure high variability, each lexical root was presented twice with alternating grammatical markings, making a total of 64 unique units for each familiarization phase. The test phase, consisting of 24 novel lexical units, was identical in all three conditions and between the two experimental groups. Learning was measured in the test phase, by the subjects' responses to the new lexical items presented.

The participants were tested individually in a research lab at the University of Bergen. When they arrived, they were asked to read and sign an informed consent form (Appendix C). They were seated in front of a laptop, asked to put on headset and follow written instructions on the computer screen (see Appendix B). E-prime 2.0 Professional (Version 2) was used for presentation of stimuli material and for collection of data. After completing the experimental tasks, the participants answered a self-rating questionnaire (Appendix A), including some auditory tasks. The questionnaire was presented on paper, and the audio files associated with the questionnaire were played on the computer. The entire experiment (the experimental tasks and the self-rating questionnaire) lasted for approximately 30 minutes.

Both researchers were present at most of the testing. Each researcher was responsible for every other participant during the time schedule. During the experimental tasks, the researchers were seated outside the test room, ready to answer eventual occurring questions or aid technical help. After each participant had completed the experiment, the responsible researcher provided them with the self-rating questionnaire and played the audio files related to the tasks. The same information and instructions were given to all participants.

Grammatical Condition. The Grammatical Condition was identical for participants in both experimental groups. The familiarization phase included 32 lexical roots with two corresponding suffixes (in total 64 unique units). The items were evenly divided between masculine and feminine lexical roots, with a correct combination of double gender markings (*-tel+ya*/*yem* for masculine roots, and *-k+oj/u* for feminine roots). All items presented in this condition were grammatically correct Russian nouns.

Substitution Condition. In the Substitution Condition, the participants were familiarized with grammatically correct nouns, as well as lexical roots with an incorrect combination of gender markings (i.e., substitution errors). Both groups were presented a total of 64 unique items, evenly split between feminine and masculine lexical roots. As presented in Table 1, the stimuli set for the Low Inconsistency group contained 25% inconsistent items (16 items) and High Inconsistency group contained 50% inconsistent items (32 items).

Table 1

Stimuli set for Low Inconsistency group (LI) and High Inconsistency group (HI) in the Substitution Condition

Lexical root	Gender markings in LI	Gender markings in HI
Dushi	-telya/telyem	-teloj*/telyem
Grabi	-teloj*/telyem	-teloj*/telyem
Obvini	-telya/telyem	-telya/telu*
Pisa	-telya/telyem	-teloj*/telyem
Rastochi	-telya/telyem	-telya/telu*
Razrushi	-teloj*/telyem	-teloj*/telyem
Potrebi	-telya/telu*	-telya/telu*
Vykljucha	-telya/telu*	-telya/telu*
Dviga	-telya/telyem	-teloj*/telyem
Khrani	-teloj*/telyem	-teloj*/telyem
Muchi	-telya/telyem	-telya/telu*
Osnova	-telya/telu*	-telya/telu*
Sluzhi	-telya/telyem	-teloj*/telyem
Smotri	-teloj*/telyem	-teloj*/telyem
Uchi	-telya/telu*	-telya/telu*
Vodi	-telya/telyem	-telya/telu*
Juboch	-kaya*/ku	-kaya*/ku
Khlopush	-koj/ku	-kaya*/ku
Makush	-kaya*/ku	-kaya*/ku
Petrush	-koj/keyem*	-koj/kayem*
Podush	-koj/ku	-koj/kayem*
Pogremush	-koj/ku	-koj/kayem*
Rodin	-koj/ku	-kaya*/ku
Soba	-koj/kayem*	-koj/kayem*
Blondin	-koj/kayem*	-koj/kayem*
Koderov	-kaya*/ku	-kaya*/ku
Konfet	-koj/ku	-kaya*/ku
Rubash	-koj/kayem*	-koj/kayem*
Skovorod	-koj/ku	-koj/kayem*
Telnjash	-koj/ku	-kaya*/ku
Rozoch	-kaya*/ku	-kaya*/ku
Vystav	-koj/ku	-koj/kayem*

Note. * = incorrect combination of gender markings (substitution errors).

Omission Condition. In the Omission Condition, the stimuli set consisted of 64 unique units, including inconsistent items of lexical roots with no gender markings (i.e., omission errors), evenly divided between feminine/masculine words. As seen in Table 2, the

Low Inconsistency group were exposed to 25% omission errors (16 items), and the High Inconsistency group to 50% omission errors (32 items).

Table 2

Stimuli set for Low Inconsistency group (LI) and High Inconsistency group (HI) in the Omission Condition

Lexical root	Gender markings in LI	Gender markings in HI
Dushi	-telya/telyem	-Ø/telyem
Grabi	-Ø/telyem	-Ø/telyem
Obvini	-telya/telyem	-telya/Ø
Pisa	-telya/telyem	-Ø/telyem
Rastochi	-telya/telyem	-telya/Ø
Razrushi	-Ø telyem	-Ø/telyem
Potrebi	-telya/Ø	-telya/Ø
Vykljucha	-telya/Ø	-telya/Ø
Dviga	-telya/telyem	-Ø/telyem
Khrani	-Ø/telyem	-Ø/telyem
Muchi	-telya/telyem	-telya/Ø
Osnova	-telya/Ø	-telya/Ø
Sluzhi	-telya/telyem	-Ø/telyem
Smotri	-Ø/telyem	-Ø/telyem
Uchi	-telya/Ø	-telya/Ø
Vodi	-telya/telyem	-telya/Ø
Juboch	-Ø/ku	-Ø/ku
Khlopush	-koj/ku	-Ø/ku
Makush	-Ø/ku	-Ø/ku
Petrush	-koj/Ø	-koj/Ø
Podush	-koj/ku	-koj/Ø
Pogremush	-koj/ku	-koj/Ø
Rodin	-koj/ku	-Ø/ku
Soba	-koj/Ø	-koj/Ø
Blondin	-koj/Ø	-koj/Ø
Koderov	-Ø/ku	-Ø/ku
Konfet	-koj/ku	-Ø ku
Rubash	-koj/Ø	-koj/Ø
Skovorod	-koj/ku	-koj/Ø
Telnjash	-koj/ku	-Ø/ ku
Rozoch	-Ø/ku	-Ø/ku
Vystav	-koj/ku	-koj/Ø

Note. * = absent gender markings (omission errors).

Test Phase. For both experimental groups, the familiarization phases were followed by an identical test phase. As presented in Table 3, the stimuli set consisted of 12 novel lexical roots, presented twice with alternating markings (one correct and one incorrect for each lexical root). The test items were divided equally between feminine/masculine words. Half of the presented items were substitution errors (6 items) and omission errors (6 items). In this phase, the participants were asked to determine each word as correct (illustrated by a smiley face) or incorrect (illustrated by a frowney face). Recall that the grammatical structures consisted of two correct masculine (*tel+ya*/*tel+yem*) and feminine (*k+oj*/*k+u*) combinations of double markings. In order to learn the correct combination of double markings, the participants had to extract the underlying structure from the familiarization phase and generalize it to the novel instances presented in the test phase.

Table 3

Stimuli set for the test phase

Lexical root	Grammatical marking	Ungrammatical marking
Deja	-telyem	-Ø
Ljubi	-telya	-teloj*
Osvezhi	-telyem	-telu*
Pokupa	-telyem	-Ø
Rodi	-telya	-Ø
Stroi	-telya	-teloj*
Brjunet	-koj	-kya*
Devush	-ku	-Ø
Maka	-ku	-Ø
Obert	-koj	-kya*
Tarel	-koj	-Ø
Verev	-ku	-kyem*

Note. * = incorrect combination of gender markings (substitution errors). Ø = absent gender markings (omission errors).

The orders in which the Stimuli Conditions were presented to the participants, were set constructing a latin-square with three conditions. A latin-square form is a technique ensuring each condition appearing one time at each ordinal position (Cozby & Bates, 2015). Our three Stimuli Conditions were set as: Grammatical Condition (1), Substitution Condition (2) and Omission Condition (3). Using a latin-square, this resulted in six potential orders of the conditions to be presented (as seen in Table 4). In the first half of the experiment, the order was set randomly in E-prime, yet the orders did not balance out in a desirable manner. We therefore decided to manipulate the order for the remaining participants, in favour of getting an approximately evenly distribution of orders among participants. This was conducted to prevent the order of exposure affecting the results.

Table 4

Orders of Stimuli Conditions presented to the participants

Order	Frequency
1-2-3	8
1-3-2	11
2-1-3	8
2-3-1	10
3-1-2	9
3-2-1	10
Total	56

Self-Rating Questionnaire. After completing the computer-based experimental tasks, participants were asked to fill out a self-rating questionnaire (see Appendix A). The questionnaire was designed to collect supplementary information on several different variables. Essentially, the participants degree of awareness, by asking about their use of rule searching strategies and their sense of achievement. It also included questions about their general language knowledge.

There are various methodical approaches for measuring awareness, including the three different measurements compared by Rebuschat, Hamrick, Riestenberg, Sachs, and Ziegler (2015): concurrent verbal reports, retrospective verbal reports and subjective measures. The two approaches of verbal reports are based on participants’ thinking aloud to verbalize their processing during training and testing (concurrent) or being interviewed about their awareness after testing (retrospective). Subjective measures of awareness are tasks including confidence ratings and source attributions, shown effective in revealing both implicit and explicit knowledge (Rebuschat et al., 2015).

Based on the approach of subjective measures of awareness, our self-rating questionnaire included source attributions and confidence ratings. For source attributions, participants were asked to deliberate the basis of their decisions in the test phases. The questions aimed to collect the participants' sense of guessing, sense of achievement and whether they applied any strategies solving the tasks, exploring the potential use of explicit learning strategies. Confidence ratings were collected as well. The participants had to report the confidence of their decisions, by rating their responses on a scale from 0 to 3 (0=*not confident*, 1= *somewhat confident,* 2= *confident,* 3= *very confident*).

During the self-rating questionnaire, 8 additional audio files were presented to the participants, linked to different types of auditory tasks. They were asked to decide whether the auditory items were correct or not (task 6 and 7), and to decide which of two different words were correct (task 10 and 11). They were also asked to decide whether they had heard the presented words earlier in the main experiment or not (task 8 and 9). Following each of these tasks, the participants had to rate their confidence (0 = *not confident;* 3 = *very confident*). This was done in order to further explore their level of confidence. All the answers from the questionnaires regarding confidence and source attributions were quantified, making them possible to use in statistical analysis.

Additionally, we wanted to obtain information about participants' general language knowledge, including their native language and the number of foreign languages they knew. In this section of the questionnaire, the question about knowledge of Russian or any other Slavic languages was repeated, emphasizing this specific exclusion criterium. Knowledge of foreign languages was specified as languages they were able to practice, including its grammatical properties. Further, they were asked to evaluate their level of knowledge in each language (0= *poor*, 1= *slightly*, 2= *good* or 3= *very good).* The number of languages rated as *good* and *very good* were quantified for use in the analysis.

Methodological Challenges

In addition to the 56 participants included in the study and described above, an additional 31 participants (15 male, 16 female) completed the experiment, resulting in a total of 87 subjects (43 male, 44 female). However, the additional 31 participants had to be excluded from the analyses and results, due to technical challenges with E-prime. In the very end of the testing done autumn 2018, we discovered that the stimuli material in the test phase was presented in a sequential order for all 28 participants in the Low Inconsistency group. This inaccuracy was caused by an error in the settings in E-prime. In addition to this, we also met some challenges in which E-prime did not play the orders of the stimuli conditions correctly, as set in the program settings. This led to a number of subjects being exposed to the same stimuli condition twice, and consequently these subjects were not exposed to all three stimuli conditions. We did not manage to detect what might have caused this inaccuracy in E-prime, and we suspect that it might be an error in the version of the program itself. This applied for three subjects in Low Inconsistency group, which were already excluded due to the first error, and for three subjects in High Inconsistency group, which we consequently had to exclude as well.

In consultation with our supervisors we decided to recruit new participants in January and February 2019, to replace the ones that were excluded due to these errors. The recruitment of new participants was conducted in the same manner as the earlier recruitment process. We additionally contacted volunteers that were put on a “backup-list” in the previous recruitment, as enough participants were sampled in the first stages of the experiment.

In the original round of recruitment, we attempted to control for age as a confounding variable, in which the participants with deviating age were evenly assigned to the experimental groups. Most participants were in their twenties and thirties, but some of them deviated in terms of age compared to the mean age of 24.4. Unfortunately, when recruiting new participants for the Low Inconsistency group in the winter of 2019, this consideration was not possible to control for. Nevertheless, all participants were above the age of 18, hence representing a sample of an adult population appropriate for our research purpose.

Analyses

In experimental research, the independent variable, also called explanatory variable, is manipulated in order to cause an anticipated change in the measured dependent variable (Frankfort-Nachmias & Nachmias, 1996). As mentioned, the basis of the current experiment included the independent variables of Experimental Group (High Inconsistency and Low Inconsistency) and Stimuli Condition (Grammatical Condition, Substitution Condition and Omission Condition). We wanted to investigate how these independent variables affected the dependent measures of Acceptance Rate (acceptance of Grammatical and Ungrammatical items).

Acceptance Rate, which is defined as significantly greater acceptance of Grammatical in contrast to Ungrammatical items, was decided to be a proper measurement of learning outcome in this experiment. This is a commonly used approach for measuring performance, based on the signal detection theory (Macmillan & Creelman, 2004). In this approach, only responses in the form of accepted items, not declined items, needs to be observed in measuring a subject's discrimination ability independent of response bias (Hochhaus & Deese, 1972). Response bias is the tendency for an overall preference for "yes" or "no" responses (Macmillan & Creelman, 2004). According to signal detection theory, acceptance of grammatical correct stimuli is termed "Hit", and acceptance of grammatical incorrect stimuli is termed "False alarm" (Kornbrot, 2006). Consequently, only the "Hit" and "False alarm" responses were used in the analysis to calculate the Acceptance Rate for each subject.

Data from the self-rating questionnaires was used to conduct correlational analyses. The independent variables from this data set were language skills and degree of awareness, chosen to investigate potential correlations with the dependent variable of Acceptance Rate. The variable of language skills was operationalized as total number of foreign languages spoken by the participant, hence excluding their first language (Norwegian). Knowledge of two other Scandinavian languages (Swedish and Danish) was also excluded, as being too closely related to Norwegian language to count as foreign. Degree of awareness as a variable, was explored by asking the participants whether they intentionally searched for regularities while listening to the words in the familiarization phases. The possible responses were quantified on a 4-point scale for use as an independent variable in the analysis (0 = *Never*, 3= *Always)*.

Various types of data require different measurement scales and hold different statistical properties. Nominal scale measurements are used for categorizing, thus unsuitable for calculating means or standard deviations (Hinton, 2004). On an interval or ratio scale however, the difference between the quantitative levels are equal in size. Additionally, ratio scales indicate the absence of the measured variable, by having an absolute zero point on the scale (Cozby & Bates, 2015). This kind of scale is typically used for measuring speed or accuracy (Hinton, 2004). In this experiment, the independent variables of Experimental Group and Stimuli Condition were treated as nominal data. The dependent variable of Acceptance Rate was measured on a ratio scale. In the correlational analyses, ratio scales were also used for the variables of language skills and degree of awareness.

The collected data was organized in Pivot tables in Excel (Version 16.25) and analysed statistically in Statistical Package for the Social Sciences (SPSS) (Version 25.0). A descriptive analysis, a mixed between- within analysis of variance (ANOVA), paired-sample *t*-tests, and correlational analyses were conducted in SPSS. Alpha (α) was set to $< .05$, which is the conventional significance level (Hinton, 2004).

Parametric tests are usually preferred, requiring interval or ratio data and an assumption of normally distributed variables (Polit & Beck, 2017). ANOVA, which was used to investigate our main hypothesis, is a parametric statistical test. Likewise, the paired-sample *t*-tests conducted to explore potential interaction effects. Additionally, parametric correlational analyses were necessary in analysing the responses from the self-rating questionnaires. Two-tailed tests were conducted, which is standard procedure when the direction of difference in the research questions is not specified in advance (Cozby & Bates, 2015).

Mixed Between-Within Analysis of Variance (ANOVA). Analysis of variance (ANOVA) is a statistical technique, allowing the researcher to compare numbers of samples and a combination of independent variables at the same time (Hinton, 2004). By using a mixed between-within subjects ANOVA in SPSS, it is possible to combine between- and within-subjects variables in a single analysis (Pallant, 2007). Including one independent (Experimental Group) and one repeated measures factor (Stimuli Condition), the design of this experiment can be classified as a two factor mixed design ANOVA, which is useful when comparing independent groups at different points across the numbers of repeated measurements (Hinton, 2004).

Between-group variance concerns the mean scores of each group, and how they differ from the grand mean. Within-group variance is the difference of individual participants scores from the group mean, hence reflecting the variability within the group (Cozby & Bates, 2015). In the current experiment, we wanted to explore the between-group effects (High Inconsistency versus Low Inconsistency) in relation to the dependent variable of Acceptance

Rate. The subjects were exposed to three different Stimuli Conditions (Grammatical, Substitution and Omission Condition). Accordingly, this created a within-group objective, as to how the different Stimuli Conditions affected the dependent variable within the groups, and further within the individual subjects.

Paired-sample *t*-test. Paired-sample *t*-tests are used for comparing a continuous variable on two occasions for one group of subjects (Pallant, 2007). We hypothesized that the Stimuli Conditions (Grammatical, Substitution and Omission Condition) would lead to significantly different learning outcomes, favouring Grammatical Condition (only grammatical correct input). Paired-sample *t*-tests were conducted in order to investigate this hypothesis, with the categorical independent variable of Stimuli Condition (Grammatical vs. Substitution vs. Omission Condition) and the continuous dependent variable of Acceptance Rate (Accepted Grammatical vs. Ungrammatical items). In these analyses, participants in both experimental groups were treated as a whole.

The overall significance of a *t*-test does not indicate the magnitude of the potential effect, which makes measurements of effect size a necessity for interpreting the results (Pallant, 2007). We applied the effect size statistic of Cohen´s *d*, commonly used for group comparisons. By calculating the *d* value, the group variance will be presented in standard deviation units of .2 (small effect), .5 (medium effect) and .8 (large effect) (Pallant, 2007).

Correlational Analyses. Correlation coefficients are measurements of relationships between variables, reflecting the strength and direction of associations (Frankfort-Nachmias & Nachmias, 1996). Pearsons *r* is a correlation coefficient which describes strength and direction of such relationships between interval or ratio scale variables (Cozby & Bates, 2015). Strength is determined by a *r* value of .5 to 1.0 as a large positive correlation, - .5 to -1.0 as a large negative correlation, and 0 suggesting no relationship between the variables (Pallant, 2007). The direction of the relationship could be positive or negative. A positive correlation is correspondingly increase or decrease in both variables, while a negative correlation is one variable increasing and the corresponding variable decreasing (Hinton, 2004).

In determining the strength and accuracy of a correlation, calculating the coefficient of determination is essential. The coefficient of determination is a calculation in which the *r* value is squared, providing a measure of shared variance between the two variables (Pallant, 2007). By doing so, we can suggest how much of the variability in one variable is explained by variability in the other. This is useful when explaining small correlations, despite of potential statistical significance (Hinton, 2004).

Correlation analyses are suitable for exploring relationships between variables. Yet, in conducting these analyses, some limitations need to be considered. First and foremost, a correlation does not indicate causality, and does not account for confounding variables. Further, in cases of curvilinear relationships, or in data sets including outliers, the correlation coefficient is likely to be underestimated (Pallant, 2007). These considerations can be accounted for, by making scatterplots prior to the analysis.

We decided to conduct the correlation coefficient Persons *r* in several correlational analyses. The objects of investigation were interrelationship between the dependent variable of Acceptance Rate in each of the three Stimuli Conditions, in correlation with the independent variables of language skills and degree of awareness. All questionnaires were collected, quantified and analysed as a whole.

Ethical aspects

There are several essential ethical considerations when conducting an experiment involving human participants. Researchers must at all times ensure and respect the participants' right to autonomy, integrity, freedom and self-determination (Den nasjonale forskningsetiske komite for samfunnsvitenskap og humaniora (NESH), 2016). Subjects in research must be sufficiently informed about all relevant aspects of the study. This includes potential benefits and risks, as well as the study's method and purpose (World Medical Association, 2013). Researchers need to obtain a consent from the participants, which must be voluntary, informed and expressed. In this sense, voluntary means that consent is given without external pressure or limitations of personal freedom of action (NESH, 2016).

In this experiment, participants were not exposed to pressure regarding attendance, due to the recruitment being based on voluntary participation. After volunteering, each participant received an e-mail, inviting them to join the project. This e-mail included all the necessary information, making sure that they would be adequately informed about all details prior to the study (Appendix D). Possibilities of withdrawal from the experiment at any time, without consequences, were also emphasized. When they arrived at the lab, they were provided with the same information sheet, and asked to sign a consent form (Appendix C) to ensure voluntary informed consent. It is important that participants don't feel pressured to sign a consent form (Creswell, 2014). To ensure that we did not apply any pressure on the participants when they were given the consent form, we stepped out of the room while they were reading and signing it. Before conducting the experimental tasks, we encouraged the participants to ask if they had any questions regarding the consent form or experiment in general.

The data material was collected and stored anonymously and there was no key linking data directly or indirectly to individual participants after the experiment was conducted. The collected data material does not consist of any direct identifying information as personal identification number, name, address or phone number. This applies for the data collected in the self-rating questionnaires as well. In addition, we have evaluated whether our data material could consist of any indirect identifying information. NSD emphasizes that this can be done by combining different information from data material, by controlling both size and composition of the sample population and by controlling for the degree of details in the gathered information (Personvernombudet for forskning, 2018). The information we obtained during the recruitment of participants (name, age, gender, language background) and all communication through e-mail and/or telephone was deleted on 01.04.2019. All of this taken into consideration, we concluded that we could carry on the experiment without applying for approval from NSD. Nevertheless, we have followed the guidelines throughout the experiment.

Another consideration is whether the sample of participants is sufficiently representative for the larger population we want to study. This is an important aspect of evidence-based practice (EBP), in generalizing evidence from an individual experiment to the real-world setting (Polit & Beck, 2017). As mentioned above, we based the inclusion of participants on the principal of “first come, first served”. The first participants to respond, and met the inclusion criteria, were included in the study. Because of this strategy of voluntary recruitment, the generalisability of the sample may have some limitations. This consideration will be discussed regarding the study's validity.

Financial compensation of research subjects is a greatly debated topic, and needs to be considered in an ethical perspective for each individual research study (Sears, 2001). As a fair appreciation for their contribution in this study, and to cover potential expenses including transportation, the subjects received a compensation of 150 NOK. To prevent any financial motivation, especially regarding recruitment at the University campus, 150 NOK were decided to be an appropriate amount.

An important ethical aspect is recognizing the properties required in the role of a researcher. Being aware of our freedom and responsibilities as researchers and conducting the study in an honest and trustworthy manner are essential aspects. This involves retaining devotion for the project, yet striving for objectivity when interpreting the results (Befring, 2002). This includes presenting results that does not support potential subjective interest as well (Creswell, 2014).

When doing research with human subjects, reflecting on the researcher-subject relation is required. Towards the participants, the researcher should always interact in a respectful manner, and strive to be gracious and polite. This includes giving permission to ask questions, or express complaints after data collection is completed (Polit & Beck, 2017). Regarding the researcher-subject relation, a few of the participants in the present study were acquaintances of us. To ensure that this would not be any ethical issue, we decided that none of us would be responsible for acquainted participants. Thus, the other researcher would have the responsibility.

Reliability

The reliability of an experiment refers to the accuracy and trustworthiness of the data material, as well as concerns for the measurements' stability and equivalence (Drageset & Ellingsen, 2009; Grønmo, 2016). In this regard, stability refers to consistency in independent data collections on the same phenomena over different time periods. Equivalence refers to consistency in independent data collections from different samples. In other words, an experiments reliability indicates to what extent the variation in the data material is caused by the methodological procedure or the implementation of the data collection (Grønmo, 2016). The data material should have high reliability regarding stability and equivalence, to ensure that the same results would have occurred if identical methodological procedures were used in another implemented data collection. In this way the experiment can be replicated, which will strengthen the findings additionally.

To ensure reliability in our study, the procedure has been thorough and systematic throughout the period of data collection. The information and instructions were identical for all of the participants, and we carefully documented everything that was done on a day to day basis. Overall, we conclude that it should be possible for other researchers to conduct a similar experiment based on our methodological procedures. That being said, it is important to emphasize that this by no means guarantees identical results in a replication of the study.

A relevant consideration concerns the inconsistent stimuli items in Omission Condition, where the audio files were edited by removing the suffixes. Six participants reported in the self-rating questionnaires that some words sounded "cut of" or "partial". Previously, similar concerns had been reported by some of the participants in our pilot study. The stimuli material was additionally edited after receiving this feedback, in an attempt to make the words sound as natural as possible. Unfortunately, some participants in the main experiment still observed this. Consequently, it must be taken into account that the omission errors might have served as clues in the process of determining correct or incorrect items. Likewise, it is a consideration in the context of self-reported attention towards word endings in general. If these auditory items originally had been recorded as omissions, it is less likely that intonation and sound quality could have potentially affected the results in the same manner.

Validity

Validity addresses whether the measured results are truly valid, or potentially reflecting alternative factors (Befring, 2002). Considering validity is an essential part of designing an experimental study. Inferences with strong internal and external validity, enhances more persuasive evidence (Polit & Beck, 2017).

Internal Validity. The internal validity of a study concerns whether it is possible to make conclusions about one variable causing changes in another variable (Cozby & Bates, 2015). To improve internal validity, the researcher has to control other potential causes and biases, making an inference that the independent variable is truthfully causing the outcome

(Polit & Beck, 2017). Identifying potential threats to the validity is crucial, making it possible to minimize these in designing the study (Creswell, 2014).

In this experiment, we attempted to minimize these threats by controlling for confounding variables and biases. Random group assignment is important in preventing selection bias, in which the results might be reflected by initial group differences (Polit & Beck, 2017). Pseudo-randomization of the participants (same number of males and females in each group), was considered appropriate in the present experiment. We wanted to prevent selection bias, yet control for the possible confounding variable of sex differences.

The potential confounding variable of age was initially managed by assigning participants with deviating age in an approximately even manner between the experimental groups. This was decided based on the mean age (24.38 years) of the participants, with only a few older participants divergent in age. Unfortunately, as previously mentioned, an error in the program settings obliged us to exclude the results of participants in Low Inconsistency group who were recruited in the first stages of the experiment. In the process of recruiting new subjects, controlling for age in the same manner was not obtainable, considering the ethical aspects of voluntary recruitment. By doing this, the standard procedure of randomization was violated, which might affect the internal validity. Pseudo-randomization was nevertheless considered a proper solution, as the experimental method enables several alternative manners of strengthening the internal validity.

To achieve a valid explanation for the observed relationships, the experimental method eliminates plausible confounding variables through experimental control (Cozby & Bates, 2015). By controlling for other potential causes than the independent variable, internal validity will be improved (Polit & Beck, 2017). Along with the confounding variable of sex difference, an identical experiment procedure was sustained collecting data from all the subjects, in order to control for differences in test situation influencing the results. The exclusion criteria of previous knowledge of Russian and other Slavic languages, was also treated as a confounding variable, and controlled for in the recruitment process. Furthermore, participants had to report previous language skills at the end of the experiment, ensuring no one conflicted this specific exclusion criteria.

The order in which the stimuli conditions is presented to the subjects, may also potentially cause biases. Ordering bias can occur when participants are exposed to different orders of intervention, because the orders themselves could affect the outcome (Polit & Beck, 2017). In the process of designing the present experiment, it was decided to construction a latin-square, preventing ordering bias caused by the order of which the stimuli conditions were presented to the subjects. In the first half of the experiment, the E-prime settings were set to random order, but we later had to manually manipulate the distribution of orders to ensure an even distribution. The final orders of presented stimuli conditions were balanced enough to suggest that ordering bias did not occur.

In experimental methods, there must be covariation between experimental group and control group to enhance internal validity (Cozby & Bates, 2015). In designing this experiment, we did not include a distinct control group. Rather, the responses in the Grammatical Condition (only grammatical correct items) were used as an individual control condition for each subject. This generates a better indication of learning outcome than a separate control group in this particular kind of research design.

In repeated measures designs, another consideration regarding the internal validity is the possibility of testing effects occurring. Testing effects are changes in the behaviour of the subject, caused by the test setting itself, which may lead to participants adjusting to the skill being tested (Cozby & Bates, 2015). In our experiment, the participants were exposed to familiarization and testing in three conditions, hence re-tested two times. This might affect the participants expectations during the experiment, which potentially could cause application of explicit strategies based on experience. Previous research has also emphasized other limitations regarding repeated measures designs. This includes a tendency for participants to accept more items overall, which has been suggested to increase after several rounds of exposure (Eidsvåg et al., 2015). In addition, repeated testing may facilitate the potential of learning occurring during the test phases (Rebuschat et al., 2015). In turn, this might limit the experiments' internal validity, and needs to be taken into consideration.

Finally, the use of a questionnaire in measuring subjective awareness, needs to be considered in terms of internal validity. Data collected from the self-reported questionnaires serve as complementary, yet not decisive, information in this type of research design. Survey as a method is potentially at risk of response bias and may result in different manners of over- or underreporting (Villar, 2008). To demonstrate, it was not sufficiently accounted for the possibility of response bias occurring, when participants were asked to specify which strategies they applied to search for input regularities. Consequently, the use of self-reported data in interpreting the results of the present experiment has its limitations.

External Validity. External validity addresses whether the results of a study can be replicated with other participants, settings and operational definitions of the variables (Cozby & Bates, 2015). A particularly important aspect is the generalizability of the study, concerning possibilities of generalizing the results from a single sample to a larger population, and from research settings to real-world settings (Polit & Beck, 2017).

Nonprobability sampling techniques, like purposive sampling of participants who meet certain criteria, may cause a selection bias. This could make it challenging to generalize the sample results to a wider population (Cozby & Bates, 2015). Using limited samples of college students might also challenge the generalizability, due to indications of student participants being a slightly more homogenous group than non-student adults (Peterson, 2001). To account for this limitation, the recruitment process was done through various channels, emphasizing the need for any volunteers above 18 years old, not college students in particular. Nevertheless, the final group of subjects included many college students between 20 and 30 years old, which is an important consideration interpreting the results. Replications of the study with other subject samples are thus necessary for generalizing to a wider population. This is a way of providing a safe guard, opposing the limited external validity of individual studies (Cozby & Bates, 2015).

Generalization also includes the extent to which the experimental situation can reflect the natural environment, and whether the laboratory setting might have an impact on the results (Frankfort-Nachmias & Nachmias, 1996). This is a major concern when generalizing tightly controlled laboratory settings to real-world settings (Polit & Beck, 2017). In the process of designing the present experiment, the contrast between an experimental setting and a natural language learning environment was a highly relevant consideration. The way in which it was accounted for, was by simulating natural variation in the stimuli material (including two types of common grammatical errors: omissions and substitutions), yet simultaneously striving to maintain the internal validity of the experiment.

Previous literature has traditionally been using artificial language stimuli in forming the theoretical foundation of statistical learning and language acquisition (Aslin & Newport, 2012; Romberg & Saffran, 2013). Artificial languages increase the internal validity of the research, by a high degree of experimental control. By using a natural language (Russian) in the present experiment, the external validity is enhanced, reflecting the real-world setting to a greater extent than artificial language stimuli. Still, there are certain challenges generalizing these kinds of measurements. Regardless of our attempts to simulate natural variability (by including omission and substitution errors) and using natural language stimuli, it is important to acknowledge the complexity of the natural language environment, in which can hardly be reflected in laboratory settings. Conducting additional studies in new settings, different times and with other groups of subjects, are ways to handle these potential issues (Creswell, 2014). Systematic reviews also assist external validity to research questions by assessing various studies in several circumstances (Polit & Beck, 2017).

Construct Validity. Construct validity concerns whether the operational definitions of variables are adequate in reflecting their proper theoretical meaning (Cozby & Bates, 2015). In our experiment, the definitions of 'statistical learning', 'inconsistency' and 'acceptance rate' require accurate operationalisations. This also enhances the experiments replicability (Grønmo, 2016).

An important concern in regard to construct validity, is operationalizing the percentages of our inconsistent stimuli items. In the present study, the levels of inconsistency are calculated by the percentage of inconsistent items in the Substitution Condition and the Omission Condition. Within each of these conditions, 25% or 50% of the presented stimuli items were inconsistent. We want to emphasize that this is one of several ways to calculate such percentages from a stimuli material like ours. In addition to the Substitution and Omission Condition, the participants were presented with 64 stimuli items in the third condition, Grammatical Condition, and 24 items in each of the three test phases. By including the items from the Grammatical Condition in the calculation (thereby including all familiarization phases) the percentages of inconsistency would be 17% (32/192 items) for the Low Inconsistency group and 33% (64/192 items) for the High Inconsistency group. A third option would be to include the test phase items in the calculation as well. By adding all three Stimuli Conditions and their following test phases, the percentage of inconsistency would be 26% (68/264 items) for the Low Inconsistency group and 38% (100/264) for the High Inconsistency group. However, we decided to calculate the percentage of inconsistency within each condition, which is most suitable for our research questions. In the present study, we aimed to assess whether the learning outcome would be affected by the different types of Stimuli Conditions, not by exposure of the stimuli material as a whole.

The aspect of construct validity was important in designing the self-report questionnaires as well. Questions regarding the independent variables of language skills and degree of awareness required proper operational definitions. In order to use these variables for statistical analyses, it was especially important to assure all participants' understanding of the questions. Prior to the experiment we conducted a small scaled pilot study (during autumn 2018), which allowed us to test, evaluate and improve the design before the larger research project. In addition to becoming comfortable in the role as researchers, it was especially important for us to get feedback on the instructions and questionnaire. After the pilot study was conducted, necessary adjustments were made, by clarifying some questions in the questionnaire which the pilot participants had trouble understanding. This included the question about previous language knowledge, in which we had to further enhance the operational definition of "knowledge". We also decided to include several comment sections in the questionnaire, making it possible to elaborate answers even further.

The questions in which we measured degree of awareness represented a challenge for the construct validity. Our wording of the relevant questions, using concepts such as 'searching for regularities' and 'applying strategies', did not include proper operational definitions. These concepts might be perceived different by individuals, yet operationalization was difficult to obtain without providing excessive information, which could further reduce its internal validity. Nevertheless, the data collected in the self-rating questionnaires offers subjective measures of awareness, which we consider a valuable supplement to the main experiment.

Concluding Remarks

The main objective of this experiment was to investigate the resilience of the statistical learning mechanisms, by exposing adults to natural language stimuli (Russian) with different percentages of inconsistencies in the input material. We wanted to explore whether the participants were able to extract the underlying structure of gender subcategories, at the levels of 25% (i.e., Low Inconsistency group) and 50% (i.e., High Inconsistency group) inconsistent input. The second objective was whether the learning outcome would be affected by different types of inconsistencies (Substitution errors and Omission errors).

Our results indicate that the learning effect of gender subcategories was resilient of inconsistent input, even at the level of 50% inconsistency. Both experimental groups (High Inconsistency and Low Inconsistency) showed significant evidence of learning, by accepting more Grammatical items than Ungrammatical items.

Previous research has suggested the learning outcome being maintained at lower levels of inconsistent input, respectively levels of 17% (Gómez & Lakusta, 2004) and 38% (Gonzales et al., 2015). Our results suggest that this level might be even higher, indicating the ability to tolerate 50% inconsistencies in the Substitution and Omission Condition. Our findings are consistent with previous literature in demonstrating that items with double gender markings facilitate learning. Due to the presence of multiple morphophonological cues, double marked nouns have been found easier to learn than single marked nouns by earlier studies (Eidsvåg et al., 2015; Gerken et al., 2005; Richardson et al., 2006).

Additionally, we found a significant learning outcome across all three Stimuli Conditions. Even after being exposed to inconsistent input with incorrect (Substitution Condition) and absence of (Omission Condition) double gender markings, both groups (High Inconsistency and Low Inconsistency) were able to learn the underlying grammatical structure and generalize to untrained items. Previous research has suggested that this learning mechanism is particularly strong in children, in contrast to adults who tend to acquire and reproduce both the consistent and inconsistent items presented to them (Hudson Kam & Newport, 2005, 2009; Pallant, 2007). Our results can thus contribute with new findings, indicating this ability to be more robust in adults than previously suggested. We found that adults were able to extract and generalize the underlying grammatical structures from different types of inconsistent input.

Overall, our results suggest a high degree of robustness in the statistical learning mechanisms in adult participants. By increasing the empirical basis on statistical learning, research in the context of language impairments can be improved as well, and further incorporated into treatment interventions (Alt, 2018; Hsu & Bishop, 2010). By contributing to the evidence base on normal learners, we also hope to enhance further research on language acquisition in individuals with language impairments.

References

Aguilar, J. M., Plante, E., & Sandoval, M. (2018). Exemplar Variability Facilitates Retention of Word Learning by Children with Specific Language Impairment. *Language, Speech, and Hearing Services in Schools, 49*(1), 72-84. doi:10.1044/2017_LSHSS-17-0031

Alt, M. (2018). Statistical Learning: How it Relates to Speech-Language Pathology.(Introduction). *Language, Speech, & Hearing Services in Schools, 49*(3 SI), 631. doi:10.1044/2018_LSHSS-STLT1-18-0040

Alt, M., Meyers, C., & Ancharski, A. (2012). Using principles of learning to inform language therapy design for children with specific language impairment. In (Vol. 47, pp. 487-498). Oxford, UK.

Antovich, D. M., & Graf Estes, K. (2018). Learning across Languages: Bilingual Experience Supports Dual Language Statistical Word Segmentation. *Developmental Science, 21*(2). doi:10.1111/desc.12548

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data.

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning. *Current Directions in Psychological Science, 21*(3), 170-176. doi:10.1177/0963721412436806

Befring, E. (2002). *Forskingsmetode, etikk og statistikk*. Oslo: Samlaget.

Braine, M. D. S. (1987). What is learned in acquiring word classes—A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 65-87). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Bulgarelli, F., Lebkuecher, A. L., & Weiss, D. J. (2018). Statistical Learning and Bilingualism. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 740-753. doi:10.1044/2018_LSHSS-STLT1-17-0139

Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 11). Cambridge, Mass: M.I.T. Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. In (Vol. 2, pp. 406-416).

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24-39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., Gremp, M. A., Walk, A. D., Bauernschmidt, A., & Pisoni, D. B. (2012). Can we enhance domain-general learning abilities to improve language function? In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition*. Berlin: De Gruyter Mouton.

Cozby, P. C., & Bates, S. C. (2015). *Methods in behavioral research* (12th ed. ed.). Boston: McGraw-Hill Education.

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.; International student ed. ed.). Los Angeles, Calif: SAGE.

de Bree, E., Verhagen, J., Kerkhoff, A., Doedens, W., & Unsworth, S. (2017). Language learning from inconsistent input: Bilingual and monolingual toddlers compared: Language learning from inconsistent input. In (pp. e1996).

Den nasjonale forskningsetiske komite for samfunnsvitenskap og humaniora (NESH). (2016). Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi. Retrieved from https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/60125_fek_retningslinjer_nesh_digital.pdf

Drageset, S., & Ellingsen, S. (2009). Forståelse av kvantitativ helseforskning - en introduksjon og oversikt. *Nordisk Tidsskrift for Helseforskning, 5*(2), 100-113. doi:10.7557/14.244

Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input Variability Facilitates Unguided Subcategory Learning in Adults. *Journal of Speech, Language, and Hearing Research, 58*(3), 826-839. doi:10.1044/2015_JSLHR-L-14-0172

Ellis, R. (2009). Implicit and Explicit Learning, Knowledge an Instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 1-3). Great Britain: Short Run Press Ltd.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321-335. doi:10.1044/1092-4388(2009/07-0189)

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822. doi:10.1073/pnas.232472899

Fiser, J., Aslin, R. N., & Lindsay, D. S. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General, 134*(4), 521-537. doi:10.1037/0096-3445.134.4.521

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences* (5th ed. ed.). London: Arnold.

Frigo, L., & McDonald, J. L. (1998). Properties of Phonological Markers That Affect the Acquisition of Gender-Like Subclasses. *Journal of Memory and Language, 39*(2), 218-245. doi:10.1006/jmla.1998.2569

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. *Cognitive Science, 33*(6), 1087-1116. doi:10.1111/j.1551-6709.2009.01041.x

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants Can Use Distributional Cues to Form Syntactic Categories. *Journal of Child Language, 32*(2), 249-268. doi:10.1017/S0305000904006786

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science, 13*(5), 431-436. doi:10.1111/1467-9280.00476

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135. doi:10.1016/S0010-0277(99)00003-7

Gómez, R. L., & Lakusta, L. (2004). A First Step in Form-Based Category Abstraction by 12-Month-Old Infants. *Developmental Science, 7*(5), 567-580. doi:10.1111/j.1467-7687.2004.00381.x

Gómez, R. L., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy, 7*(2), 183-206. doi:10.1207/s15327078in0702_4

Gonzales, K., Gerken, L., & Gómez, R. L. (2015). Does hearing two dialects at different times help infants learn dialect-specific rules? *Cognition, 140*, 60-71. doi:10.1016/j.cognition.2015.03.015

Gonzales, K., Gerken, L., & Gómez, R. L. (2018). How who is talking matters as much as what they say to infant language learners. *Cognitive Psychology, 106*, 1-20. doi:10.1016/j.cogpsych.2018.04.003

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can Infants Map Meaning to Newly Segmented Words?: Statistical Segmentation and Word Learning. *Psychological Science, 18*(3), 254-260. doi:10.1111/j.1467-9280.2007.01885.x

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The Effects of Variation on Learning Word Order Rules by Adults with and without Language-Based Learning Disabilities. *Journal of Communication Disorders, 39*(2), 158-170. doi:10.1016/j.jcomdis.2005.11.004

Grønmo, S. (2016). *Samfunnsvitenskapelige metoder* (2. utg. ed.). Bergen: Fagbokforl.

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical Word Learning in Children with Autism Spectrum Disorder and Specific Language Impairment. *Journal of Child Psychology and Psychiatry, 58*(11), 1251-1263. doi:10.1111/jcpp.12734

Hall, J., Owen Van Horne, A. J., McGregor, K. K., & Farmer, T. A. (2018). Individual and Developmental Differences in Distributional Learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 694-709. doi:10.1044/2018_LSHSS-STLT1-17-0134

Hamrick, P., & Rebuschat, P. (2012). How implicit are statistical learning? . In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365-382). Berlin: De Gruyter Mouton.

Hay, J. F., & Lany, J. (2012). Sensitivity to statistical information begets learning in early language development. In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 91-118). Berlin: De Gruyter, Inc.

Hinton, P. R. (2004). *Statistics explained* (2nd ed. ed.). London: Routledge.

Hochhaus, L., & Deese, J. (1972). A table for the calculation of d' and BETA. *Psychological Bulletin, 77*(5), 375-376. doi:10.1037/h0032459

Horvath, S., McDermott, E., Reilly, K., & Arunachalam, S. (2018). Acquisition of Verb Meaning from Syntactic Distribution in Preschoolers with Autism Spectrum Disorder. *Language, Speech, and Hearing Services in Schools, 49*(3 SI), 668-680. doi:10.1044/2018_LSHSS-STLT1-17-0126

Hsu, H. J., & Bishop, D. V. M. (2010). Grammatical Difficulties in Children with Specific Language Impairment: Is Learning Deficient? *Human Development, 53*(5), 264-277. doi:10.1159/000321289

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development, 1*(2), 151-195. doi:10.1080/15475441.2005.9684215

Hudson Kam, C. L., & Newport, E. L. (2009). Getting It Right by Getting It Wrong: When Learners Change Languages. *Cognitive Psychology, 59*(1), 30-66. doi:10.1016/j.cogpsych.2009.01.001

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116. doi:10.1016/0163-6383(95)90012-8

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual Statistical Learning in Infancy: Evidence for a Domain General Learning Mechanism. *Cognition, 83*(2), B35-B42. doi:10.1016/S0010-0277(02)00004-5

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbjørnsen, A. E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian *. *Bilingualism, 13*(4), 513-523. doi:10.1017/S1366728910000039

Kornbrot, D. (2006). Signal detection theory, the approach of choice: Model-based and distribution-free measures and evaluation. *Perception & Psychophysics, 68*(3), 393-414. doi:10.3758/BF03193685

Kovács, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science (New York, N.Y.), 325*(5940), 611. doi:10.1126/science.1173947

Lian, A. (2017). Statistical learning and developmental language impairments. *Statistical learning and developmental language impairments, 4*(e9).

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science, 11*(1), 122-134. doi:10.1111/j.1467-7687.2007.00653.x

Mintz, T. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*(5), 678-686. doi:10.3758/BF03196424

Mintz, T. H. (2003). Frequent Frames as a Cue for Grammatical Categories in Child Directed Speech. *Cognition, 90*(1), 91-117. doi:10.1016/S0010-0277(03)00140-9

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The Link between Statistical Segmentation and Word Learning in Adults. *Cognition, 108*(1), 271-280. doi:10.1016/j.cognition.2008.02.003

Montgomery, J. W., & Leonard, L. B. (1998). Real-Time Inflectional Processing by Children with Specific Language Impairment: Effects of Phonetic Substance. *Journal of Speech, Language, and Hearing Research, 41*(6), 1432-1443. doi:10.1044/jslhr.4106.1432

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162. doi:10.1016/S0010-0285(03)00128-2

Niemeier, S. (1999). A cognitive view on bilingualism and "bilingual" teaching and learning. *Journal of English Studies*(1), 165-186.

Owen Van Horne, A. J., Curran, M., Larson, C., & Fey, M. E. (2018). Effects of a Complexity-Based Approach on Generalization of Past Tense "-ed" and Related Morphemes. *Language, Speech, and Hearing Services in Schools, 49*(3S), 681-693. doi:10.1044/2018_LSHSS-STLT1-17-0142

Pallant, J. (2007). *SPSS survival manual : a step by step guide to data analysing using SPSS for Windows* (3rd ed. ed.). Maidenhead: McGraw-Hill ; Open University Press.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development, 80*(3), 674-685. doi:10.1111/j.1467-8624.2009.01290.x

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science (New York, N.Y.), 298*(5593), 604. doi:10.1126/science.1072901

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238. doi:10.1016/j.tics.2006.03.006

Personvernombudet for forskning. (2018). Norsk Samfunnsvitenskapelig Datatjeneste (NSD). Retrieved from http://www.nsd.uib.no/personvernombud/meld_prosjekt/index.html

Peterson, Robert A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research, 28*(3), 450-461. doi:10.1086/323732

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*(5), 453-462. doi:10.1016/S0021-9924(02)00094-1

Plante, E., & Gómez, R. L. (2018). Learning without Trying: The Clinical Relevance of Statistical Learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 710-722. doi:10.1044/2018_LSHSS-STLT1-17-0131

Polit, D. F., & Beck, C. T. (2017). *Nursing Research : generating and assessing evidence for nursing practice* (10th ed. ed.). Philadelphia: Wolters Kluwer.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855-863.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating Measures of Awareness: A Contribution to the Debate on Learning without Awareness. *Studies in Second Language Acquisition, 37*(2), 299-334. doi:10.1017/S0272263115000145

Rebuschat, P., & Williams, J. N. (2012). Introduction: Statistical learning and language acquisition. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 1-12). Berlin: De Gruyter Mouton.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From Shared Contexts to Syntactic Categories: The Role of Distributional Information in Learning Linguistic Form-Classes. *Cognitive Psychology, 66*(1), 30-54. doi:10.1016/j.cogpsych.2012.09.001

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language, 97*, 17-29. doi:10.1016/j.jml.2017.07.006

Richardson, J., Harris, L., Plante, E., & Gerken, L. (2006). Subcategory Learning in Normal and Language Learning-Disabled Adults: How Much Information Do They Need? *Journal of Speech, Language, and Hearing Research, 49*(6), 1257-1266. doi:10.1044/1092-4388(2006/090)

Romberg, A. R., & Saffran, J. R. (2013). All Together Now: Concurrent Learning of Multiple Structures in an Artificial Language. *Cognitive Science, 37*(7), 1290-1320. doi:10.1111/cogs.12050

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical Pattern Learning by Human Infants and Cotton-Top Tamarin Monkeys. *Cognition, 107*(2), 479-500. doi:10.1016/j.cognition.2007.10.010

Saffran, J. R. (2003). Statistical Language Learning. *Current Directions in Psychological Science, 12*(4), 110-114. doi:10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. (1996a). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926. doi:10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*(4), 606-621. doi:10.1006/jmla.1996.0032

Saffran, J. R., & Thiessen, E. D. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology, 39*(3), 484-494. doi:10.1037/0012-1649.39.3.484

Sandoval, M., Gonzales, K., & Gómez, R. (2012). The road to word class acquisition is paved with distributional and sound cues. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 145-169). Berlin: De Gruyter, Inc.

Sandoval, M., Patterson, D., Dai, H., Plante, E., & Vance, C. J. (2017). Neural Correlates of Morphology Acquisition through a Statistical Learning Paradigm.(Report). *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.01234

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Sears, J. M. (2001). Payment of Research Subjects: A Broader Perspective. *The American Journal of Bioethics, 1*(2), 66-67. doi:10.1162/152651601300169185

Shanks, D. (2003). Attention and awareness in «implicit» sequence learning. In L. Jimenez (Ed.), *Attention and Implicit Learning* (pp. 11-42). Philadelphia, PA: John Benjamins.

Torkildsen, J. V. K., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar Variability Facilitates Rapid Learning of an Otherwise Unlearnable Grammar by

Individuals with Language-Based Learning Disability. *Journal of Speech, Language, and Hearing Research, 56*(2), 618-629. doi:10.1044/1092-4388(2012/11-0125)

Toro, J. M., Trobalon, J. B., Sebastián-Gallés, N., & Mackintosh, N. (2005). Effects of Backward Speech and Speaker Variability in Language Discrimination by Rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(1), 95-100. doi:10.1037/0097-7403.31.1.95

Villar, A. (2008). Response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 752-753). Thousand Oaks: United States of America, California, Thousand Oaks: Sage Publications, Inc.

Wade, T., & Gillespie, D. (2011). *A comprehensive Russian grammar* (3rd ed. rev. and updated by David Gillespie. ed.). Malden, Mass: Wiley-Blackwell.

Wang, F. H., Mintz, T. H., Greene, R. L., & Benjamin, A. S. (2018). Learning Nonadjacent Dependencies Embedded in Sentences of an Artificial Language: When Learning Breaks Down. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 604-614. doi:10.1037/xlm0000483

Wilson, R., Gerken, L., & Nicol, J. (2002). Syntactic category learning in a second language. Retrieved from https://search.proquest.com/docview/304802212?accountid=8579

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and Processing Verb Argument Structure: Distributional Learning in a Miniature Language. *Cognitive Psychology, 56*(3), 165-209. doi:10.1016/j.cogpsych.2007.04.002

World Medical Association. (2013). WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Running head: LEARNING FROM INCONSISTENT INPUT

The Effects of Inconsistent Input in Grammar Learning: How Much are Adult Learners Able to Tolerate?

Andrea Eide and Elisabeth Sæterøy Skeistrand

Master's Programme in Health Sciences, Speech-Language Pathology

The Faculty of Psychology

Department of Biological and Medical Psychology

The University of Bergen

Spring 2019

Abstract

We explored the robustness of statistical learning mechanisms by presenting noun gender subcategories in an unfamiliar language (Russian) to adult language learners. We investigated how much inconsistent input they were able to tolerate, and whether different types of inconsistencies would affect the learning. A total of 56 adults (28 male and 28 female) participated in the experiment, evenly allocated in two experimental groups (Low Inconsistency and High Inconsistency). Participants in both groups were familiarized with masculine and feminine Russian nouns, followed by a test phase. Familiarization and testing were completed three times in total, with three different conditions of familiarization stimuli. For two of the familiarization conditions, the stimuli included inconsistent items (substitution errors or omission errors). The remaining condition included only grammatically correct items. During familiarization, participants in the Low Inconsistency group heard 25% inconsistent stimuli, while the High Inconsistency group heard 50% inconsistent stimuli. During the test phase, participants were asked to distinguish between grammatically correct and incorrect words for generalization. Participants in both experimental groups (Low Inconsistency and High Inconsistency) showed a statistically significant learning outcome. Our findings did not reveal any difference in learning between the three stimuli conditions. The results reveal that high levels of input inconsistency (50 %) still facilitate learning of gender subcategories. The participants were able to extract the underlying grammatical structure, even when exposed to input including items with alternating and absent gender markings.

Key words: Language acquisition, Statistical learning, Inconsistency, Natural language stimuli

The Effects of Inconsistent Input in Grammar Learning: How Much are Adult Learners Able to Tolerate?

Statistical learning is in broad terms described as a mechanism for pattern recognition present in infants, children and adults (Aslin & Newport, 2012). The term was introduced in an influential study by Saffran et al. (1996a), demonstrating that 8-month-old infants are able to use transitional properties between syllables to discover word boundaries in a continuous speech stream from an artificial language. Since then, a lot of research has been conducted investigating both infants, children and adult learners' ability to detect such statistical properties in their language input. Statistical learning has traditionally been studied in the context of language acquisition, but is intertwined with the implicit learning research tradition, exploring the incidental domain-general learning mechanisms in humans (Perruchet & Pacton, 2006). To define learning as implicit, the learner must be without awareness and use of explicit hypothesis-driven strategies (Cleeremans et al., 1998). The distinction between implicit and explicit learning is the object of several methodological approaches, such as measuring the subjects' degree of awareness (Rebuschat et al., 2015).

Statistical learning is described as a robust mechanism (Alt, 2018), and a growing body of research have provided evidence that infants, children and adults are able to detect statistical properties to solve different linguistic tasks. This includes detection of word boundaries (Pelucchi et al., 2009; Saffran et al., 1996a; Saffran et al., 1996b), phonological and phonetic regularities (e.g., Maye, Weiss, & Aslin, 2008; Saffran & Thiessen, 2003) and importantly for the present study, acquisition of grammatical categories and subcategories (e.g., Eidsvåg et al., 2015; Gerken et al., 2005; Mintz, 2003; Reeder et al., 2013, 2017; Richardson et al., 2006; Wonnacott, Newport, & Tanenhaus, 2008).

In our everyday lives, we are exposed to complex and varied language input. As talkers, we have varied pronunciation, vocabulary and grammar, due to differences in gender, dialects and proficiency (Gonzales et al., 2018). Language learners are exposed to both consistent and inconsistent language input every day. Thus, a central issue is how much inconsistent input the learner can tolerate, and still be able to extract the statistical properties of the language structure. Gómez and Lakusta (2004) explored this in a series of experiments, by adding inconsistent items in an artificial grammar system. In one of their experiments, they investigated 12-month-old infants' ability to learn grammatical categories. A total of 48 infants (50/50 male female) were divided into two groups (17% grammatical inconsistencies or 33% grammatical inconsistencies) and exposed to one of two artificial training languages (one consisting of *aX* and *bY* pairings, and the other of *aY* and *bX* pairings). They created grammatical inconsistencies by inserting elements from the other language, which violated the grammatical rules of the training language (Gómez & Lakusta, 2004). Their findings showed that infants were able to learn the artificial grammar structure with 17% inconsistent input, but infants failed to learn at 33% inconsistency. These results suggested that the limit of how much inconsistencies the infant learner can tolerate before learning is disrupted, might be a place in between these percentages (Gómez & Lakusta, 2004).

Gonzales et al. (2015) however, found that infants might be able to tolerate a higher degree of inconsistencies in their input when the different dialects are presented in separate streams. In this study, 12-month-old infants were exposed to two artificial language streams, representing different dialects. The "pure stream" followed the abstract grammar rule *aX bY*. In the "mixed stream" the sentences randomly violated the pure stream's rules in which any *a*- or *b*- element could be paired with any *X*- or *Y*-word. In total, 108/288 sentences violated the pure stream's rules, which constitutes a total of 38% inconsistent elements. Importantly, all *X*-words were monosyllabic and the *Y*-words were disyllabic, and thereby distinguishable by syllable length. In three experiments, they found that infants were able to acquire the pure stream's grammatical rules even when the auditory streams contained 38% of inconsistent elements. Yet, the streams were required to be heard separately and not randomly interleaved. Infants in this study showed evidence of learning even without any perceptual cues (e.g., different voices) (Gonzales et al., 2015).

In a later study, Gonzales and collaborators (2018) exposed 12-month-olds to two artificial speech streams, representing different language models. They asked whether the presence of different perceptual cues distinguishing the streams would affect the infants learning. The stimuli material used in this experiment was highly similar to the one in their previous study (Gonzales et al., 2015). However, in two of their experiments (experiment 2 and 3) the percentage of inconsistent items in the mixed stream was 25%. In these experiments the infants heard the two speech streams played interleaved by different voices (experimental condition) or by the same voice (control condition). Infants in both experiments demonstrated learning of the pure stream's rules, only when the streams were presented by different talkers (Gonzales et al., 2018).

In a natural language environment, grammatical inconsistencies, such as errors in articulation or inflection of words, are frequently heard in children (particularly with language impairments) and second language learners (Bulgarelli et al., 2018; Montgomery & Leonard, 1998). In the present study, we used substitution and omission errors as inconsistent stimuli items. Substitution errors are created by lexical roots with incorrect grammatical markers. Omission errors are lexical roots without any grammatical markings.

Hudson Kam and Newport (2005, 2009) have conducted several experiments aimed to investigate how different grammatical inconsistencies in artificial languages are learned, both with children and adult participants. The inconsistent items included omission errors (Hudson Kam & Newport, 2005) and substitution errors (Hudson Kam & Newport, 2009). In their first study, they showed that adult learners tend to acquire and reproduce the same amount of inconsistencies that were presented in the training stimuli. The amount ranged from

45% to 100% exposure of determiners, thus the remaining percentages were words with no determiners (omission errors). However, children regularized the inconsistent forms most of the time (90%), and thus acquired and reproduced the consistent structure (Hudson Kam & Newport, 2005). In their later study (Hudson Kam & Newport, 2009), the participants were exposed to sentences containing the main determiner (*ka* or *po*) 60% of the time. The remaining 40% of the time, substitution determiners (e.g., *te*, *meg*, *li'* etc.) occurred in varied frequencies across conditions (ranging from two substitution determiners to sixteen different substitution determiners). Their findings demonstrated that the adult participants' learning of the main determiners increased with increased variation in the substitution determiners. When the variation of different substitution determiner forms was at its highest (sixteen different), adults produced the main determiners almost 90% of the time. This result may indicate that adult learners benefit from the presence of high variability, as suggested in several previous studies (e.g., Eidsvåg et al., 2015). When comparing children (mean age 5 years, 10.6 months) and adult learners (mean age 20 years, 2.25 months) they found that children regularized the inconsistent forms, as in their previous study, while adult learners only showed evidence of such regularization of the most complex inconsistencies (Hudson Kam & Newport, 2009).

The presented results strengthen the evidence that statistical learning is a robust mechanism, but indicates that it might be particularly robust in young children (Hudson Kam & Newport, 2009). In the present study, we aimed to further investigate adults' learning from inconsistent input. We wanted to assess whether the previous suggested tendencies in infants and children also apply for adult language learners. The participants in the present study are Norwegians, which in general are exposed to several foreign languages in school and throughout their everyday life. By some definitions, many of the participants can be considered as functional bilinguals (for explanation, see Niemeier, 1999).

In recent years, researchers have begun to explore the impact of bilingualism on statistical learning. Bilingual infants have to acquire multiple linguistic systems at the same time, as they are exposed to a smaller amount of input from the different languages (Bulgarelli et al., 2018). Research findings suggest that early bilingualism may serve as an advantage in tracking multiple sets of statistics (Antovich & Graf Estes, 2018; Kovács & Mehler, 2009). This bilingual advantage may extend throughout adulthood, but these suggestions are still inconclusive (for a detailed review on statistical learning and bilingualism, see Bulgarelli et al., 2018).

In regard to inconsistencies in the input, de Bree et al. (2017) conducted a study comparing monolingual and bilingual toddlers' (24-month-olds) learning of non-adjacent dependencies. They wanted to explore whether the presence of inconsistent items would affect learning. Their findings show that neither group learned the non-adjacent dependencies when the training input only contained grammatical correct items. Interestingly, bilingual toddlers were able to detect the structure when 14% of the input were inconsistent items (de Bree et al., 2017). These results indicate that bilingual infants may have an increased sensitivity to their presented input, due to their experience with multiple language structures. This might enhance their learning when they are presented with structures that are inconsistent (de Bree et al., 2017). Gonzales et al. (2018) also suggested a bilingual advantage in their experiment, in which greater exposure to a second language was linked to a novelty preference during the testing. In Experiment 2, the infants had no more than 10 hours of weekly exposure to a second language, while infants in Experiment 3 had a greater exposure to a second language.

As mentioned, earlier studies have documented successful learning in infants with a percentage of 17, 25 and 38 inconsistent input, when the input is presented in specific conditions (Gómez & Lakusta, 2004; Gonzales et al., 2015, 2018). In the present study, we have chosen the levels of 25% and 50 % inconsistency to assess how much adult participants can tolerate. Based on the previous findings, we expect the learning effects to be maintained at 25%, but disrupted at 50%. Thereby, our main hypothesis is that the participants exposed to 25% inconsistent items will demonstrate a significantly greater learning outcome than the participants exposed to 50% inconsistency.

Based on findings from Hudson Kam and Newport (2005, 2009), we expect that the presence of inconsistent items may decrease adult learners' ability to detect the consistent grammatical structures. Accordingly, we hypothesize that the presence of both substitution errors and omission errors will affect the learning outcome. We expect a better performance of both groups in the test phase following the Grammatical Condition, due to the absence of ungrammatical input.

Method

Participants

Fifty-six adults were included in the present experiment, twenty-eight male (mean age of 25.07, *SD* 6.7) and twenty-eight female (mean age 23.68, *SD* 5.08). All participants were 18 years old or above and had Norwegian as their native language, as required by the inclusion criteria. Four participants had additional native languages, including English (n=1), Spanish (n=2) and Turkish (n=1). Forty-nine participants reported knowledge of two foreign languages or more. The remaining participants (n=7) reported knowledge of one foreign language (English). The exclusion criteria for participating in the experiment were self-reported hearing loss, language deficits, developmental or acquired neurological disorders. Additionally, the participants were required to have no prior knowledge of Russian or other related Slavic languages.

We allocated the participants evenly into two groups: one receiving 25% inconsistent items (Low Inconsistency Group) and one group receiving 50% inconsistent items (High

Inconsistency Group). The allocation was based on pseudo-randomization. To control for potential sex differences, we arranged an equal number of males and females in each group.

Participants in the Low Inconsistency group were between the ages of nineteen and thirty-three (*M* 23.0, *SD* 3.11), and in High Inconsistency group between eighteen and fifty-six (*M* 25.75, *SD* 7.62). Two participants in each group had an additional native language. The general language knowledge (number of foreign languages spoken by the participants, in addition to Scandinavian languages similar to Norwegian), were ranging from 1 and 6 in the Low Inconsistency group (mean 2.46, *SD* 1.04) and 1 to 5 in the High Inconsistency group (mean 2.18, *SD* .905). There was no significant difference between the groups, regarding age or language knowledge.

Research design

The present experiment has a 2x3 experimental design, in which the two experimental groups (High Inconsistency and Low Inconsistency) were exposed to three Stimuli Conditions during familiarization: Grammatical Condition, Substitution Condition and Omission Condition. Each condition consisted of a familiarization phase of auditory input, and a test phase. During the Grammatical Condition, the familiarization input consisted of correct combinations of lexical roots and gender marking suffixes, resulting in only grammatically correct exemplars and serving as a control condition. During the Substitution Condition, the stimuli contained grammatically correct items, but also items with incorrect combinations of gender suffixes, thus acting as inconsistent stimuli. The Omission Condition included grammatically correct items, in addition to inconsistent items where the suffixes had been removed. Percentage of inconsistency in these phases differed between the groups, with respectively 25% (Low Inconsistency) and 50% (High Inconsistency) incorrect input. The test phases were identical in all conditions.

Stimuli material

The participants were exposed to auditory Russian input, consisting of feminine and masculine lexical roots with a subset of possible suffixes. Similar stimuli material has previously been used by Gerken et al. (2005), Richardson et al. (2006) and Eidsvåg et al. (2015). In our experiment, we decided to only use items with double gender markings, which have shown to be easier to learn, relative to single marked items (Eidsvåg et al., 2015; Gerken et al., 2005; Richardson et al., 2006).

In the present stimuli material, both the double gender markings -*tel+ya* and -*tel+yem* could be applied to the masculine lexical roots. For the feminine lexical roots, the correct markings applied were -*k+oj* and -*k+u*. As an example, the feminine lexical root *Blondin* could be inflected by both of the associated gender markings (e.g. *Blondin-k+oj* or *Blondin-k+u*), forming a grammatically correct noun. Consequently, lexical roots with non-associated or absent suffixes resulted in grammatically incorrect items, in which the inconsistent items for the Substitution Condition and the Omission Condition was constructed.

The audio material consisted of one audio file for each stimuli item, recorded by native Russian speakers. The auditory input in the familiarization phases were recorded by one female talker and one male talker, while the input in the test phase was presented by a novel female talker. Prior to the present experiment, the audio files were edited in GoldWave Version 6.35 (GoldWave Inc, 2019), filtering out background noise to achieve the best sound quality possible. Some of the audio files were additionally edited by identifying and removing the gender markings for use as inconsistent items in the Omission Condition. The suffixes were removed from these items, remaining single units of lexical roots that contained the same spectre qualities of the root words for all items.

Familiarization Phase. The familiarization stimuli for each condition included 32 lexical roots, with or without gender markings, in total 64 unique words. In all conditions, the stimuli items were presented in a randomized order. In the Grammatical Condition, the items were evenly divided between masculine and feminine lexical roots, with a correct combination of double gender markings (*-tel+ya/yem* for masculine roots, and *-k+oj/u* for feminine roots). Table 1 presents the stimuli set for the familiarization phases in Substitution Condition and Omission Condition for both experimental groups. In these conditions, the Low Inconsistency group was exposed to 25% inconsistent items (16 items), while High Inconsistency group was presented 50% inconsistent items (32 items).

Table 1 here

Test Phase. Stimuli for the test phase, as seen in Table 2, consisted of 12 new lexical roots, presented with two different suffixes (a total of 24 unique words) in a randomized order. Twelve of the presented items were grammatical, twelve were ungrammatical (six were substitution errors, six were omission errors).

Table 2 here

Self-Rating Questionnaire. A self-rating questionnaire (see Appendix A) was included in order to assess participants' degree of awareness. Based on the approach of subjective measures of awareness (Rebuschat et al., 2015), the questions included source attributions and confidence ratings. Participants were asked to elaborate potential use of strategies, and to rate their confidence on a scale (0= *not confident*, 1= *somewhat confident*, 2= *confident*, 3= *very confident*).

We also included some auditory tasks in order to explore the levels of confidence even further. Six audio files were presented during these tasks. Two were lexical roots from the familiarization stimuli material (Task 6 and 9), two were lexical roots from the test stimuli (Task 10 and 11), and the remaining two were not previously heard (Task 7 and 8).

Additionally, the self-reported questionnaire was used to collect information about the participants' general language skills (native language(s) and foreign language knowledge). The variable of Language skills was operationalized as the total number of foreign languages spoken by the participant, excluding their first language (Norwegian) and closely related languages (Swedish and Danish).

Procedure

All participants were tested individually in a research lab at the University of Bergen, Norway. After signing an informed consent form, they were provided with headphones and asked to follow the written instructions on a computer screen. E-prime 2.0 Professional (Version 2) was used to present the audio files, as well as collecting the data. Necessary instructions were presented on the screen prior to the familiarization and test phases (see Appendix B), yet the researchers were always nearby to clarify potential questions, or for assistance if needed.

During the experimental tasks, participants were exposed to three stimuli conditions (Grammatical Condition, Substitution Condition and Omission Condition). The orders in which the three conditions were presented, were balanced across participants using a latin-square design. This resulted in six potential orders of stimuli conditions. After exposure to each familiarization phase, the participants were presented with the test phase and asked to determine whether each word they heard was correct or incorrect. The responses were given by clicking on a "smiley face" (for correct words) or a "frowney face"(for incorrect word) on the computer screen.

The participants were asked to answer a self-rating questionnaire presented on paper after completing the experimental tasks (see Appendix A). As mentioned above, the questionnaire included some auditory tasks in addition to the written questions.

The entire experiment lasted for approximately 30 minutes. After completing the experiment, participants received a compensation of NOK 150 (approximately USD 20), to cover expenses associated with participation.

The data material was collected and stored anonymously and did not consist of any identifying personal information. There was no key linking data directly or indirectly to individual participants after the experiment was conducted. This taken into consideration, the experiment was conducted without applying for approval from Norwegian Social Science Data Services (NSD). Nevertheless, the NSD guidelines were pursued throughout the experiment (Personvernombudet for forskning, 2018).

Analysis

The main objective in this experiment was to investigate whether the Low Inconsistency group resulted in significantly greater learning of the subset of Russian grammar, compared to High Inconsistency group. Further, we explored whether different types of inconsistent input (substitution errors and omission errors) had a statistically significant effect on the learning outcome.

Learning was measured by Acceptance Rate, which is operationalized by a significantly higher acceptance of Grammatical items versus acceptance of Ungrammatical items. According to the signal detection theory model, the acceptance of grammatical correct stimuli is termed “Hit”, and acceptance of grammatical incorrect stimuli is referred to as “False alarm” (Kornbrot, 2006).

Additionally, data from the self-rating questionnaires was used to investigate potential correlations between the independent variables (i.e., degree of awareness and language skills), and the dependent variable (i.e., Acceptance Rate).

Descriptive analysis, mixed between- within analysis of variance (ANOVA), paired-sample *t*-tests and correlational analyses were conducted in order to analyse the collected data. Two-tailed analyses were performed, with alpha (α) set to $< .05$.

Results

Acceptance Rate

The data was analysed conducting a 2 x 2 x 3 mixed design ANOVA. Experimental Group (High Inconsistency vs. Low Inconsistency) served as the between-group factor, Acceptance Rate (accepted Grammatical vs. accepted Ungrammatical items) and Stimuli Condition (Grammatical Condition vs. Substitution Condition vs. Omission Condition) were set as within-group factors.

Descriptive statistics for Acceptance Rate for all Stimuli Conditions are presented in Table 3. The main effect of Acceptance Rate (Accepted Grammatical vs. Ungrammatical items) was statistically significant, $F_{(1, 54)} = 245.38$, $p<.05$, $\mu_p^2 = .82$, revealing significant effects of learning. The between-subject effect of Experimental Group (High Inconsistency vs. Low Inconsistency) showed no significant results, $F_{(1, 54)} = .137$, *ns*, $\mu_p^2 = .03$, hence no overall difference in learning between the groups.

Table 3 here

The main effect of Stimuli Condition (Grammatical vs. Substitution vs. Omission Condition) was significant, $F_{(2, 108)} = 3.27$, $p<.05$, $\mu_p^2 = .06$, as well as the interaction effect of Stimuli Condition (Grammatical vs. Substitution vs. Omission Condition) x Acceptance Rate (Accepted Grammatical vs. Ungrammatical items), $F_{(2, 108)} = 6.49$, $p<.05$, $\mu_p^2 = .001$.

Paired-sample *t*-tests were conducted in order to investigate this interaction effect. The *t*-tests revealed higher acceptance of Grammatical items than Ungrammatical items in the Grammatical Condition, $t(55) = -13.43$, $p < .0005$, $d = 2.156$, as well as the Substitution

Condition, $t(55) = -11.80$ $p < .0005$, $d = 2.324$, and the Omission Condition, $t(55) = -6.92$, $p < .0005$, $d = 1.187$. The effect size was large ($d > .8$) in all Stimuli Conditions (Cohen, 1988). The results demonstrate a statistically significant learning outcome in all of the three Stimuli Conditions, suggesting that the significant interaction effect may be due to mean scores unrelated to the experiment. Mean numbers of accepted items in all Stimuli Conditions (Grammatical Condition, Substitution Condition and Omission Condition) for both groups (High Inconsistency and Low Inconsistency) are shown in Figure 1.

Figure 1 here

Self-Rating Questionnaires

To analyse the data from the self-rating questionnaires, the correlation coefficient Pearson's *r* was measured conducting correlational analyses. The object of investigation was the interrelationship of Acceptance Rate in each of the Stimuli Conditions, in correlation with the independent variables of self-reported degree of awareness and language skills.

Degree of awareness yielded a positive correlation ($r = .41$, $p < .05$) with accepted Grammatical items in the Omission Condition. The analysis of language skills and Acceptance Rate resulted in a significant negative correlation for number of accepted Grammatical items in the Substitution Condition, $r = -.34$ ($p = .010$). No other significant correlations were found.

In the self-rating questionnaires, the participants were asked whether they searched for regularities in the presented words. All participants reported searching for regularities. Further, when asked if they applied any strategies while performing the experimental task, 28 participants answered that they often (n= 26) or always (n=2) did, while the remaining half did not (n= 11), or only occasionally (n=17).

In order to collect data on source attributions, participants were asked to elaborate (*Which strategies/rules did you apply*?). Twenty-seven participants, (11 in Low Inconsistency Group, 16 in High Inconsistency Group), reported searching for regularities in the suffixes. One participant in the High Inconsistency Group correctly recalled the suffix *-koj*, while others (in both Low and High Inconsistency Group) incorrectly reported recognizing the suffixes *-kom, -em, - a, -o* and *-j.* Six of the participants reported noticing some of the word endings sounding “cut of” or “partial”. These words were consequently perceived as “unnatural”, hence determined as incorrect in the test phase. Five of these six participants were in the Low Inconsistency group. Eleven participants reported that their attention was drawn to phonological patterns as well. One participant in the Low Inconsistency group reported noticing the number of syllables, while several others mentioned being attentive of phonemes or pronunciation. Sixteen participants did not specify any applied strategies.

The questions regarding their use of strategies included confidence ratings. The majority of participants (69.6%) reported low confidence in whether the strategies applied were correct or not. In the additional auditory tasks, the mean score of confidence was 1.24 (1= *somewhat confident,* 2= *confident*).

Discussion

The purpose of our experiment was to investigate the robustness of statistical language learning, by exploring the effects of different amounts and types of inconsistent input in a simplified Russian grammar system. Our main findings were that overall, both experimental groups (Low Inconsistency group and High Inconsistency group) revealed significant learning effects. We also found a significant learning effect for all Stimuli Conditions (Grammatical Condition, Substitution Condition and Omission Condition) when analysed separately.

Participants in both groups accepted more Grammatical items than Ungrammatical items, hence showing a significant learning outcome. These results indicate no statistically significant effect between the two groups, rejecting our main hypothesis that lower percentage of inconsistent items (25%) in the presented stimuli material, facilitates greater learning effects than a higher percentage (50%). The findings suggest a clearly robust learning mechanism in adult language learners, as the participants were able to extract the underlying grammatical system even at the level of 50% inconsistent input.

Previous literature on infants have indicated evidence of learning being maintained only at lower levels of inconsistencies, respectively the levels of 17 % (Gómez & Lakusta, 2004), 25% (Gonzales et al., 2018) and 38% (Gonzales et al., 2015). Other studies have found that learners can tolerate higher levels of inconsistent input to some degree, but this learning effect have been suggested to require additional contextual cues in the input (Gebhart, Aslin, & Newport, 2009; Gonzales et al., 2018). Our stimuli material did not include any supplementary contextual cues such as voice characteristics, previously being found essential for extracting grammatical structures in infant language learners (Gonzales et al., 2018). Nevertheless, the stimuli items contained double markings, which is shown to enhance learning relative to single marked items, due to the presence of multiple morphophonological cues (Eidsvåg et al., 2015; Gerken et al., 2005; Richardson et al., 2006).

Some methodological challenges have been disclosed in these previous studies. When high levels of an inconsistent structure are presented to the participants, the predominant structure of consistent elements will be less salient. As pointed out by Gómez and Lakusta (2004), it remains unclear whether the inconsistent input disrupted the learning effect, or whether the infants learned both of the structures simultaneously. By measuring learning as a significantly higher acceptance of grammatical versus ungrammatical items, our results show that the participants did not learn the alternative structures (substitution errors and omission errors). The participants were not aware of what was defined as Grammatical and Ungrammatical. Still, the participants exposed to 50% inconsistent items in the Substitution and Omission Condition (same amount of Substitution and Omission errors as Grammatical items), accepted more Grammatical items than Ungrammatical items during testing.

The ability to extract the underlying structure, even when exposed to 50% inconsistent input, is an unexpected finding. However, there are several manners of calculating the levels of inconsistencies. The items presented in the Grammatical Condition (only Grammatically correct) and test phases (50% Grammatical and 50% Ungrammatical) were not included in the present calculation of percentages. By including all presented stimuli items in the experiment as a whole, the percentages of inconsistency changes. During the whole experiment, all participants were exposed to 264 items. The overall percentages of inconsistent items presented would thus be 26% (68/264 words) in the Low Inconsistency group, and 38% (100/264) in the High Inconsistency group. Percentage of inconsistencies in the High Inconsistency group would thereby decrease, as well as the difference in percentage between the two groups. This might serve as a possible explanation to why both groups managed to learn the grammatical pattern.

Participants in both groups (Low Inconsistency and High Inconsistency) were able to extract the underlying grammatical structure and generalize to untrained items, even after being exposed to different types of inconsistent stimuli. The learning outcome was significant in all Stimuli Conditions, hence rejecting our second hypothesis that the learning effect would be significantly greater after exposure to the Grammatical Condition. Previous research has indicated this tendency in children, who learned the regular structures of inconsistent input (Hudson Kam & Newport, 2005, 2009). Adult learners however, tended to acquire the language structures as presented to them, including inconsistencies (Hudson Kam & Newport, 2005, 2009). In contrast, the findings in the present experiment suggests that adults have the ability to extract and generalize underlying grammatical structures from inconsistent input as well. By using adult participants, these findings can contribute to the broader field of literature on statistical learning mechanisms, which previously have focused mainly on children and infants (Gerken et al., 2005; Gonzales et al., 2015, 2018).

Self-Reported Degree of Awareness

Based on data collected from the self-rating questionnaires, a significant positive correlation was found for accepted Grammatical items in the Omission Condition, in relationship to degree of awareness. No correlations were found in the other Stimuli Conditions. In this regard, the omission stimuli may raise a potential issue. As previously mentioned, the omission items were made prior to this experiment, by removing the suffixes from some of the audio files. Six participants reported that some of words from the presented audio material sounded cut or incomplete. This raises an issue, as to whether the quality of these stimuli items potentially attracted attention towards the word endings. Nevertheless, these concerns remain speculative. A possible solution for this limitation is to record original omission error items in future research. Hence, the rest of the auditory material should be recorded by the same new voices as well, to ensure overall consistency.

All participants reported searching for regularities during the experiment. Still, the majority was not confident about the correctness of their eventual use of strategies. When asked to elaborate, forty participants were able to explicitly report several kinds of strategies. Despite proposals of potential explicit strategies, the learning process of the underlying grammatical structure might remain unaware (Ellis, 2009). Based on these subjective reports of awareness and confidence ratings, we cannot conclude whether the participants in truth applied any explicit strategies in the present learning process.

We also want to emphasize that data collected from self-reported questionnaires serve as complementary, yet not decisive, information in this type of research design. Survey as a method is potentially at risk of response bias and may result in different manners of over- or underreporting (Villar, 2008). An important consideration is the fact that the responses were reported in retrospect, after the experiment was conducted. If explicit strategies were applied, we cannot conclude whether the participants applied these strategies in all rounds of testing. Measuring awareness during the experiment might in turn increase the participants' awareness, potentially leading them to apply explicit strategies such as hypothesis testing. Conducting studies where these limitations are accounted for, are issues to be handled for future research.

Participants' Language Knowledge

Another consideration worth mentioning, is the general language skills of the participants in our study. The analysis of language skills did not correlate with the responses in all Stimuli Conditions, indicating that the participants' individual language knowledge did not affect the results. However, as part of the Norwegian educational system, Norwegians are in general exposed to multiple foreign languages. Recall that almost all of the participants reported knowledge of two foreign languages or more in the self-rating questionnaire (n=49). Four participants (two in each group) had an additional native language as well. Overall, this demonstrates that the participants have been exposed to multiple foreign languages, and that they generally can be considered as a population sample with a high degree of language skills. Only the four participants with an additional native language are considered bilingual from birth. However, the sample population in general can be considered as functional bilinguals, as previously mentioned, since the participants were (or had recently been) students (Niemeier, 1999).

The participants general high exposure to foreign languages might be a possible explanation to why our results do not correspond with previous findings in regard to inconsistent input. Previous research regarding the impact of bilingualism on statistical learning, indicate that exposure to a second language may be advantageous, and possibly serve as "protection" for inconsistent items in the input (de Bree et al., 2017; Gonzales et al., 2018). Thereby, we cannot rule out the possibility that our participants' knowledge of foreign languages might have affected the results. We encourage future research to investigate the possible link between previous language exposure and resilience to inconsistent items. This could be done by exploring the robustness of statistical learning with similar methods in bilingual sample populations. Another possibility could be to compare experimental groups of monolinguals with little previous exposure to foreign languages, with groups of monolinguals with greater exposure to foreign languages.

Conclusions

The results of the present experiment indicate that the learning effect of gender subcategories is resilient to inconsistent input, even at the level of 50% inconsistency. After the familiarization, participants in both experimental groups (Low Inconsistency and High Inconsistency) showed evidence of learning in which they had an overall higher acceptance rate for Grammatical items than Ungrammatical items at test.

Additionally, a significant learning outcome was found across different stimuli conditions of inconsistent input in adult language learners. The subjects demonstrated an ability to distinguish correct Russian nouns, from items with incorrect or absent gender marking.

Acknowledgements

The experiment was funded by the Department of Biological and Medical Psychology at the University of Bergen. We would like to thank Dr. Elena Plante at the University of Arizona for contributing with additional audio files to our stimuli material, and Bjørn Sætrevik for programming the experiment in E-prime.

References

Aguilar, J. M., Plante, E., & Sandoval, M. (2018). Exemplar Variability Facilitates Retention of Word Learning by Children with Specific Language Impairment. *Language, Speech, and Hearing Services in Schools, 49*(1), 72-84. doi:10.1044/2017_LSHSS-17-0031

Alt, M. (2018). Statistical Learning: How it Relates to Speech-Language Pathology.(Introduction). *Language, Speech, & Hearing Services in Schools, 49*(3 SI), 631. doi:10.1044/2018_LSHSS-STLT1-18-0040

Alt, M., Meyers, C., & Ancharski, A. (2012). Using principles of learning to inform language therapy design for children with specific language impairment. In (Vol. 47, pp. 487-498). Oxford, UK.

Antovich, D. M., & Graf Estes, K. (2018). Learning across Languages: Bilingual Experience Supports Dual Language Statistical Word Segmentation. *Developmental Science, 21*(2). doi:10.1111/desc.12548

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data.

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning. *Current Directions in Psychological Science, 21*(3), 170-176. doi:10.1177/0963721412436806

Befring, E. (2002). *Forskingsmetode, etikk og statistikk*. Oslo: Samlaget.

Braine, M. D. S. (1987). What is learned in acquiring word classes—A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 65-87). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Brooks, P., Braine, M., Catalano, L., Brody, R., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of Memory and Language, 32*(1), 76.

Bulgarelli, F., Lebkuecher, A. L., & Weiss, D. J. (2018). Statistical Learning and Bilingualism. *Language, Speech, and Hearing Services in Schools, 49*(3S), 740-753. doi:10.1044/2018_LSHSS-STLT1-17-0139

Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 11). Cambridge, Mass: M.I.T. Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. In (Vol. 2, pp. 406-416).

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Conway, C. M., & Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 24-39. doi:10.1037/0278-7393.31.1.24

Conway, C. M., Gremp, M. A., Walk, A. D., Bauernschmidt, A., & Pisoni, D. B. (2012). Can we enhance domain-general learning abilities to improve language function? In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition*. Berlin: De Gruyter Mouton.

Cozby, P. C., & Bates, S. C. (2015). *Methods in behavioral research* (12th ed. ed.). Boston: McGraw-Hill Education.

Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed.; International student ed. ed.). Los Angeles, Calif: SAGE.

de Bree, E., Verhagen, J., Kerkhoff, A., Doedens, W., & Unsworth, S. (2017). Language learning from inconsistent input: Bilingual and monolingual toddlers compared: Language learning from inconsistent input. In (pp. e1996).

Den nasjonale forskningsetiske komite for samfunnsvitenskap og humaniora (NESH). (2016). Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi.

Retrieved from https://www.etikkom.no/globalassets/documents/publikasjoner-som-pdf/60125_fek_retningslinjer_nesh_digital.pdf

Drageset, S., & Ellingsen, S. (2009). Forståelse av kvantitativ helseforskning - en introduksjon og oversikt. *Nordisk Tidsskrift for Helseforskning, 5*(2), 100-113. doi:10.7557/14.244

Eidsvåg, S. S., Austad, M., Plante, E., & Asbjørnsen, A. E. (2015). Input Variability Facilitates Unguided Subcategory Learning in Adults. *Journal of Speech, Language, and Hearing Research, 58*(3), 826-839. doi:10.1044/2015_JSLHR-L-14-0172

Ellis, R. (2009). Implicit and Explicit Learning, Knowledge an Instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 1-3). Great Britain: Short Run Press Ltd.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical Learning in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 52*(2), 321-335. doi:10.1044/1092-4388(2009/07-0189)

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15822. doi:10.1073/pnas.232472899

Fiser, J., Aslin, R. N., & Lindsay, D. S. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General, 134*(4), 521-537. doi:10.1037/0096-3445.134.4.521

Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences* (5th ed. ed.). London: Arnold.

Frigo, L., & McDonald, J. L. (1998). Properties of Phonological Markers That Affect the Acquisition of Gender-Like Subclasses. *Journal of Memory and Language, 39*(2), 218-245. doi:10.1006/jmla.1998.2569

Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing Structures in Midstream: Learning Along the Statistical Garden Path. *Cognitive Science, 33*(6), 1087-1116. doi:10.1111/j.1551-6709.2009.01041.x

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants Can Use Distributional Cues to Form Syntactic Categories. *Journal of Child Language, 32*(2), 249-268. doi:10.1017/S0305000904006786

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science, 13*(5), 431-436. doi:10.1111/1467-9280.00476

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135. doi:10.1016/S0010-0277(99)00003-7

Gómez, R. L., & Lakusta, L. (2004). A First Step in Form-Based Category Abstraction by 12-Month-Old Infants. *Developmental Science, 7*(5), 567-580. doi:10.1111/j.1467-7687.2004.00381.x

Gómez, R. L., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy, 7*(2), 183-206. doi:10.1207/s15327078in0702_4

Gonzales, K., Gerken, L., & Gómez, R. L. (2015). Does hearing two dialects at different times help infants learn dialect-specific rules? *Cognition, 140*, 60-71. doi:10.1016/j.cognition.2015.03.015

Gonzales, K., Gerken, L., & Gómez, R. L. (2018). How who is talking matters as much as what they say to infant language learners. *Cognitive Psychology, 106*, 1-20. doi:10.1016/j.cogpsych.2018.04.003

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can Infants Map Meaning to Newly Segmented Words?: Statistical Segmentation and Word Learning. *Psychological Science, 18*(3), 254-260. doi:10.1111/j.1467-9280.2007.01885.x

Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The Effects of Variation on Learning Word Order Rules by Adults with and without Language-Based Learning Disabilities. *Journal of Communication Disorders, 39*(2), 158-170. doi:10.1016/j.jcomdis.2005.11.004

Grønmo, S. (2016). *Samfunnsvitenskapelige metoder* (2. utg. ed.). Bergen: Fagbokforl.

Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical Word Learning in Children with Autism Spectrum Disorder and Specific Language Impairment. *Journal of Child Psychology and Psychiatry, 58*(11), 1251-1263. doi:10.1111/jcpp.12734

Hall, J., Owen Van Horne, A. J., McGregor, K. K., & Farmer, T. A. (2018). Individual and Developmental Differences in Distributional Learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 694-709. doi:10.1044/2018_LSHSS-STLT1-17-0134

Hamrick, P., & Rebuschat, P. (2012). How implicit are statistical learning? . In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365-382). Berlin: De Gruyter Mouton.

Hay, J. F., & Lany, J. (2012). Sensitivity to statistical information begets learning in early language development. In P. Rebuschat & J. N. Williams (Eds.), *Statistical Learning and Language Acquisition* (pp. 91-118). Berlin: De Gruyter, Inc.

Hinton, P. R. (2004). *Statistics explained* (2nd ed. ed.). London: Routledge.

Hochhaus, L., & Deese, J. (1972). A table for the calculation of d' and BETA. *Psychological Bulletin, 77*(5), 375-376. doi:10.1037/h0032459

Horvath, S., McDermott, E., Reilly, K., & Arunachalam, S. (2018). Acquisition of Verb Meaning from Syntactic Distribution in Preschoolers with Autism Spectrum Disorder. *Language, Speech, and Hearing Services in Schools, 49*(3 SI), 668-680. doi:10.1044/2018_LSHSS-STLT1-17-0126

Hsu, H. J., & Bishop, D. V. M. (2010). Grammatical Difficulties in Children with Specific Language Impairment: Is Learning Deficient? *Human Development, 53*(5), 264-277. doi:10.1159/000321289

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development, 1*(2), 151-195. doi:10.1080/15475441.2005.9684215

Hudson Kam, C. L., & Newport, E. L. (2009). Getting It Right by Getting It Wrong: When Learners Change Languages. *Cognitive Psychology, 59*(1), 30-66. doi:10.1016/j.cogpsych.2009.01.001

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116. doi:10.1016/0163-6383(95)90012-8

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual Statistical Learning in Infancy: Evidence for a Domain General Learning Mechanism. *Cognition, 83*(2), B35-B42. doi:10.1016/S0010-0277(02)00004-5

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbjørnsen, A. E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian *. *Bilingualism, 13*(4), 513-523. doi:10.1017/S1366728910000039

Kornbrot, D. (2006). Signal detection theory, the approach of choice: Model-based and distribution-free measures and evaluation. *Perception & Psychophysics, 68*(3), 393-414. doi:10.3758/BF03193685

Kovács, A. M., & Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. *Science (New York, N.Y.), 325*(5940), 611. doi:10.1126/science.1173947

Kusumawardhani, P. (2017). The analysis of errors of omission in english narrative composition made by EFL students. In (pp. 84).

Lian, A. (2017). Statistical learning and developmental language impairments. *Statistical learning and developmental language impairments, 4*(e9).

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd ed. ed.): Psychology Press.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science, 11*(1), 122-134. doi:10.1111/j.1467-7687.2007.00653.x

Mintz, T. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*(5), 678-686. doi:10.3758/BF03196424

Mintz, T. H. (2003). Frequent Frames as a Cue for Grammatical Categories in Child Directed Speech. *Cognition, 90*(1), 91-117. doi:10.1016/S0010-0277(03)00140-9

Mirman, D., Magnuson, J. S., Estes, K. G., & Dixon, J. A. (2008). The Link between Statistical Segmentation and Word Learning in Adults. *Cognition, 108*(1), 271-280. doi:10.1016/j.cognition.2008.02.003

Montgomery, J. W., & Leonard, L. B. (1998). Real-Time Inflectional Processing by Children with Specific Language Impairment: Effects of Phonetic Substance. *Journal of Speech, Language, and Hearing Research, 41*(6), 1432-1443. doi:10.1044/jslhr.4106.1432

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162. doi:10.1016/S0010-0285(03)00128-2

Niemeier, S. (1999). A cognitive view on bilingualism and "bilingual" teaching and learning. *Journal of English Studies*(1), 165-186.

Owen Van Horne, A. J., Curran, M., Larson, C., & Fey, M. E. (2018). Effects of a Complexity-Based Approach on Generalization of Past Tense "-ed" and Related Morphemes. *Language, Speech, and Hearing Services in Schools, 49*(3S), 681-693. doi:10.1044/2018_LSHSS-STLT1-17-0142

Pallant, J. (2007). *SPSS survival manual : a step by step guide to data analysing using SPSS for Windows* (3rd ed. ed.). Maidenhead: McGraw-Hill ; Open University Press.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development, 80*(3), 674-685. doi:10.1111/j.1467-8624.2009.01290.x

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science (New York, N.Y.), 298*(5593), 604. doi:10.1126/science.1072901

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences, 10*(5), 233-238. doi:10.1016/j.tics.2006.03.006

Personvernombudet for forskning. (2018). Norsk Samfunnsvitenskapelig Datatjeneste (NSD). Retrieved from http://www.nsd.uib.no/personvernombud/meld_prosjekt/index.html

Peterson, Robert A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research, 28*(3), 450-461. doi:10.1086/323732

Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*(5), 453-462. doi:10.1016/S0021-9924(02)00094-1

Plante, E., & Gómez, R. L. (2018). Learning without Trying: The Clinical Relevance of Statistical Learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 710-722. doi:10.1044/2018_LSHSS-STLT1-17-0131

Polit, D. F., & Beck, C. T. (2017). *Nursing Research : generating and assessing evidence for nursing practice* (10th ed. ed.). Philadelphia: Wolters Kluwer.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855-863.

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating Measures of Awareness: A Contribution to the Debate on Learning without Awareness. *Studies in Second Language Acquisition, 37*(2), 299-334. doi:10.1017/S0272263115000145

Rebuschat, P., & Williams, J. N. (2012). Introduction: Statistical learning and language acquisition. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 1-12). Berlin: De Gruyter Mouton.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From Shared Contexts to Syntactic Categories: The Role of Distributional Information in Learning Linguistic Form-Classes. *Cognitive Psychology, 66*(1), 30-54. doi:10.1016/j.cogpsych.2012.09.001

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language, 97*, 17-29. doi:10.1016/j.jml.2017.07.006

Richardson, J., Harris, L., Plante, E., & Gerken, L. (2006). Subcategory Learning in Normal and Language Learning-Disabled Adults: How Much Information Do They Need?

Journal of Speech, Language, and Hearing Research, 49(6), 1257-1266. doi:10.1044/1092-4388(2006/090)

Romberg, A. R., & Saffran, J. R. (2013). All Together Now: Concurrent Learning of Multiple Structures in an Artificial Language. *Cognitive Science, 37*(7), 1290-1320. doi:10.1111/cogs.12050

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical Pattern Learning by Human Infants and Cotton-Top Tamarin Monkeys. *Cognition, 107*(2), 479-500. doi:10.1016/j.cognition.2007.10.010

Saffran, J. R. (2003). Statistical Language Learning. *Current Directions in Psychological Science, 12*(4), 110-114. doi:10.1111/1467-8721.01243

Saffran, J. R., Aslin, R. N., & Newport, E. (1996a). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926. doi:10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*(4), 606-621. doi:10.1006/jmla.1996.0032

Saffran, J. R., & Thiessen, E. D. (2003). Pattern Induction by Infant Language Learners. *Developmental Psychology, 39*(3), 484-494. doi:10.1037/0012-1649.39.3.484

Sandoval, M., Gonzales, K., & Gómez, R. (2012). The road to word class acquisition is paved with distributional and sound cues. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 145-169). Berlin: De Gruyter, Inc.

Sandoval, M., Patterson, D., Dai, H., Plante, E., & Vance, C. J. (2017). Neural Correlates of Morphology Acquisition through a Statistical Learning Paradigm.(Report). *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.01234

Santos, F. H. (2012). Substitution Errors in Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (Vol. 24,2). Boston, MA: Boston, MA: Springer US.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Sears, J. M. (2001). Payment of Research Subjects: A Broader Perspective. *The American Journal of Bioethics, 1*(2), 66-67. doi:10.1162/152651601300169185

Shanks, D. (2003). Attention and awareness in «implicit» sequence learning. In L. Jimenez (Ed.), *Attention and Implicit Learning* (pp. 11-42). Philadelphia, PA: John Benjamins.

Torkildsen, J. V. K., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar Variability Facilitates Rapid Learning of an Otherwise Unlearnable Grammar by Individuals with Language-Based Learning Disability. *Journal of Speech, Language, and Hearing Research, 56*(2), 618-629. doi:10.1044/1092-4388(2012/11-0125)

Toro, J. M., Trobalon, J. B., Sebastián-Gallés, N., & Mackintosh, N. (2005). Effects of Backward Speech and Speaker Variability in Language Discrimination by Rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(1), 95-100. doi:10.1037/0097-7403.31.1.95

Villar, A. (2008). Response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods* (pp. 752-753). Thousand Oaks: United States of America, California, Thousand Oaks: Sage Publications, Inc.

Wade, T., & Gillespie, D. (2011). *A comprehensive Russian grammar* (3rd ed. rev. and updated by David Gillespie. ed.). Malden, Mass: Wiley-Blackwell.

Wang, F. H., Mintz, T. H., Greene, R. L., & Benjamin, A. S. (2018). Learning Nonadjacent Dependencies Embedded in Sentences of an Artificial Language: When Learning Breaks Down. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 604-614. doi:10.1037/xlm0000483

Wilson, R., Gerken, L., & Nicol, J. (2002). Syntactic category learning in a second language. Retrieved from https://search.proquest.com/docview/304802212?accountid=8579

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and Processing Verb Argument Structure: Distributional Learning in a Miniature Language. *Cognitive Psychology, 56*(3), 165-209. doi:10.1016/j.cogpsych.2007.04.002

World Medical Association. (2013). WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Table 1

Stimuli set for Low Inconsistency group (LI) and High Inconsistency group (HI) in the Substitution Condition and Omission Condition

Lexical root	Gender markings in LI	Gender markings in HI
Dushi	-telya/telyem	-teloj*/telyem
Grabi	-teloj*/telyem	-teloj*/telyem
Obvini	-telya/telyem	-telya/telu*
Pisa	-telya/telyem	-teloj*/telyem
Rastochi	-telya/telyem	-telya/telu*
Razrushi	-teloj*/telyem	-teloj*/telyem
Potrebi	-telya/telu*	-telya/telu*
Vykljucha	-telya/telu*	-telya/telu*
Dviga	-telya/telyem	-teloj*/telyem
Khrani	-teloj*/telyem	-teloj*/telyem
Muchi	-telya/telyem	-telya/telu*
Osnova	-telya/telu*	-telya/telu*
Sluzhi	-telya/telyem	-teloj*/telyem
Smotri	-teloj*/telyem	-teloj*/telyem
Uchi	-telya/telu*	-telya/telu*
Vodi	-telya/telyem	-telya/telu*
Juboch	-kaya*/ku	-kaya*/ku
Khlopush	-koj/ku	-kaya*/ku
Makush	-kaya*/ku	-kaya*/ku
Petrush	-koj/keyem*	-koj/kayem*
Podush	-koj/ku	-koj/kayem*
Pogremush	-koj/ku	-koj/kayem*
Rodin	-koj/ku	-kaya*/ku
Soba	-koj/kayem*	-koj/kayem*
Blondin	-koj/kayem*	-koj/kayem*
Koderov	-kaya*/ku	-kaya*/ku

Konfet	-koj/ku	-kaya*/ku
Rubash	-koj/kayem*	-koj/kayem*
Skovorod	-koj/ku	-koj/kayem*
Telnjash	-koj/ku	-kaya*/ku
Rozoch	-kaya*/ku	-kaya*/ku
Vystav	-koj/ku	-koj/kayem*

Note. * = In Substitution Condition, these suffixes consisted of an incorrect combination of gender markings (i.e., substitution errors), as presented in the table. In Omission Condition, these suffixes were absent (i.e., omission errors).

Table 2

Stimuli set for the test phase

Lexical root	Grammatical marking	Ungrammatical marking
Deja	-telyem	-∅
Ljubi	-telya	-teloj*
Osvezhi	-telyem	-telu*
Pokupa	-telyem	-∅
Rodi	-telya	-∅
Stroi	-telya	-teloj*
Brjunet	-koj	-kya*
Devush	-ku	-∅
Maka	-ku	-∅
Obert	-koj	-kya*
Tarel	-koj	-∅
Verev	-ku	-kyem*

Note. * = incorrect combination of gender markings (substitution errors). ∅ = absent gender markings (omission errors).

Table 3

Descriptive statistics for Acceptance Rate (accepted Grammatical vs. Ungrammatical items) for Low Inconsistency (LI) (n= 28) and High Inconsistency (HI) (n=28) in Grammatical Condition (GC), Substitution Condition (SC) and Omission Condition (OC)

Acceptance Rate	Mean LI	Std.Dev. LI	Mean HI	Std.Dev. HI
Grammatical GC	8.9	(2.2)	9.1	(1.6)
Ungrammatical GC	5.1	(1.8)	5.0	(1.6)
Grammatical SC	8.7	(1.8)	9.4	(2.0)
Ungrammatical SC	5.3	(1.5)	5.0	(1.1)
Grammatical OC	8.6	(2.1)	8.9	(2.2)
Ungrammatical OC	6.2	(2.4)	6.3	(2.0)

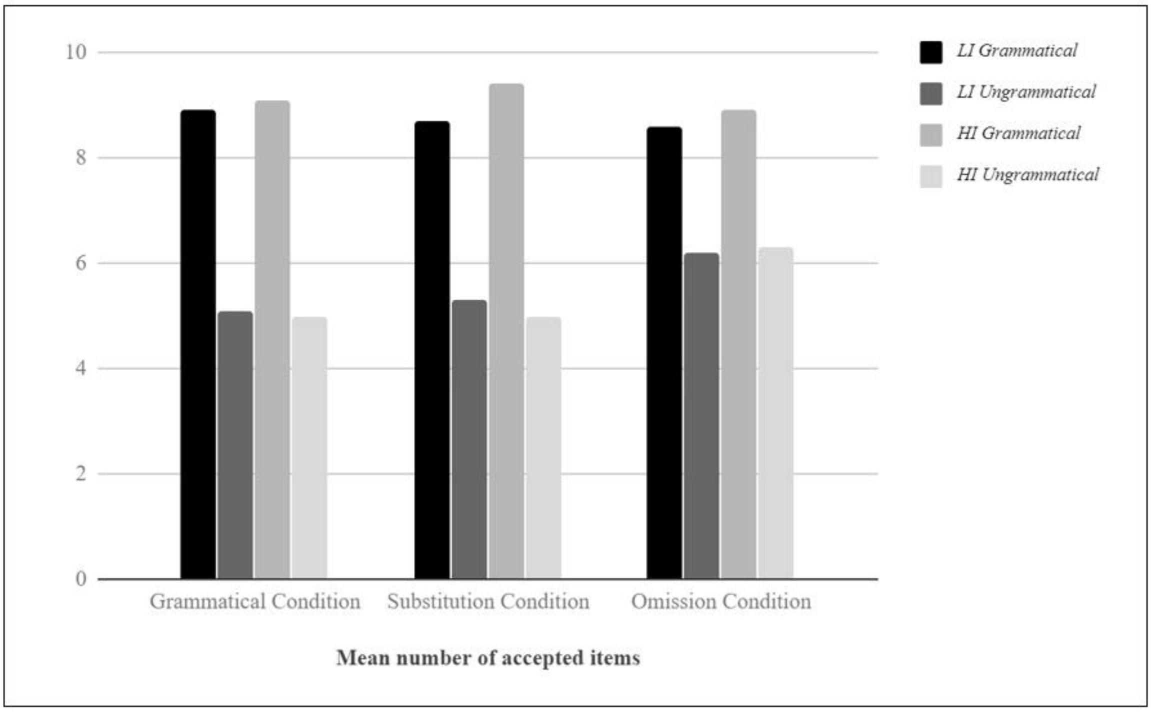


Figure 1. Mean number of accepted items (Grammatical vs. Ungrammatical items) for Low Inconsistency group (LI) (n=28) and High Inconsistency group (HI) (n=28) in all Stimuli Conditions (Grammatical, Substitution and Omission Condition).

Appendix A

Selvrapporteringsskjema

(Self-Rating Questionnaire)

Spørsmål om de databaserte oppgavene (ring rundt det svaralternativet som passer)
(Questions about the experimental tasks)

1. Tre ganger i løpet av forsøket ble du bedt om å bedømme om ordene du hørte var korrekte eller ukorrekte. Hvilken runde tror du at du mestret best?

   (You were asked to determine which words were correct and incorrect three times during the experiment. In which round did you feel the greatest achievement?)

   Runde 1 *(Round 1)* Runde 2 *(Round 2)* Runde 3 *(Round 3)* Vet ikke *(I don't know)*

   Eventuelt hvorfor? *(If so, why?)*

   ____

   ____

2. Følte du at avgjørelsene dine var tilfeldige?
   (Did you feel that your decisions were incidental?)

   Nei *(No)* Av og til *(Sometimes)* Ofte *(Often)* Hele tiden *(Always)*

3. Lette du etter regelmessigheter/mønstre i ordene du hørte?
   (Did you search for any regularities in the presented words?)

   Nei *(No)* Litt *(Somewhat)* I stor grad *(To a large extent)* I svært stor grad *(To a very large extent)*

4. Fulgte du noen regler/mønstre da du svarte på oppgavene?
   (Did you apply any strategies/rules during the tasks?)

Nei	Litt	I stor grad	I svært stor grad
(No)	*(Somewhat)*	*(To a large extent)*	*(To a very large extent)*

Eventuelt hvilke? *(If so, which?)*

____

____

5. Tror du at de reglene/mønstrene du eventuelt fulgte var korrekte?
(Do you think that the strategies/rules you applied were correct?)

Ja	Nei	Vet ikke
(Yes)	*(No)*	*(I don't know)*

Konkrete spørsmål (ring rundt det svaralternativet som passer)
(Specific questions)

6. Er dette ordet korrekt?
(Is this word correct?)

Ja	Nei
(Yes)	*(No)*

Hvor sikker er du?
(How confident are you?)

Usikker	Litt usikker	Sikker	Svært sikker
(Not confident)	*(Somewhat confident)*	*(Confident)*	*(Very confident)*

7. Er dette ordet korrekt?
(Is this word correct?)

Ja	Nei
(Yes)	*(No)*

Hvor sikker er du?

(How confident are you?)

Usikker Litt usikker Sikker Svært sikker

(Not confident) (Somewhat confident) (Confident) (Very confident)

8. Var dette et av ordene du lyttet til?

(Was this one of the words you were previously presented with?)

Ja Nei

(Yes) (No)

Hvor sikker er du?

(How confident are you?)

Usikker Litt usikker Sikker Svært sikker

(Not confident) (Somewhat confident) (Confident) (Very confident)

9. Var dette et av ordene du lyttet til?

(Was this one of the words you were previously presented with?)

Ja Nei

(Yes) (No)

Hvor sikker er du?

(How confident are you?)

Usikker Litt usikker Sikker Svært sikker

(Not confident) (Somewhat confident) (Confident) (Very confident)

10. Nå skal du få høre to ord. Hvilket av disse ordene er korrekt?

(You will now be presented two words. Which one is correct?)

Ord 1 Ord 2

(Word 1) (Word 2)

Hvor sikker er du?

(How confident are you?)

Usikker Litt usikker Sikker Svært sikker

(Not confident) (Somewhat confident) (Confident) (Very confident)

11. Nå skal du få høre to ord. Hvilket av disse ordene er korrekt?

(You will now be presented two words. Which one is correct?)

Ord 1 Ord 2

(Word 1) (Word 2)

Hvor sikker er du?

(How confident are you?)

Usikker Litt usikker Sikker Svært sikker

(Not confident) (Somewhat confident) (Confident) (Very confident)

Generelle språkkunnskaper (ring rundt det svaralternativet som passer)

(General language knowledge)

12. Er norsk ditt morsmål?

(Is Norwegian your native language?)

Ja Nei

(Yes) (No)

Eventuelle andre morsmål:

(Do you have any additional native languages?)

____

13. Nevn hvilke andre fremmedspråk du har kjennskap til, og bedøm graden av ferdighetene dine i parentes bak det enkelte språket (svært lite, litt, godt, svært godt). Med kjennskap menes språk du kan bruke, inkludert grammatikk.

(Which foreign languages do you know? Additionally, evaluate your skills in each language (poor, little, good, very good). Only include languages you can practice, including its grammatical properties.

____

____

____

14. Har du kjennskap til grammatikk i russisk eller andre slaviske språk?

(Do you have knowledge of Russian grammar or other Slavic languages?)

Ja Nei

(Yes) *(No)*

Appendix B

Initial instructions presented to the participants on the computer screen.

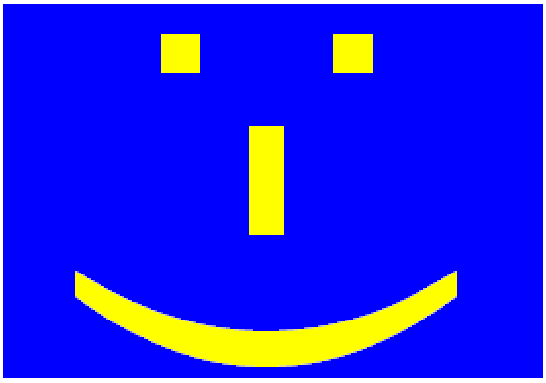
(Welcome to this experiment!
We will now investigate something called"implicit language learning". Implicit language learning is when we unintentionally learn something about a language just by listening to it.
Press SPACE to continue)

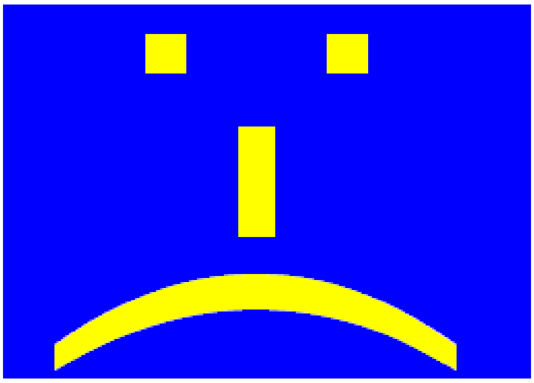
Instructions given before each of the familiarization phases.

(You will now hear some words.
You have to listen carefully, because afterwards you will be given some tasks.
Press SPACE when you are ready.)

Instructions given before each of the test phases.

(You will now hear some new words.
Decide whether the words you hear are correct or incorrect. Press "smiley" for correct, and "frowney" for incorrect.
Press SPACE when you are ready.)





(That's all!
Thank you for participating!)

Appendix C

Consent form signed by the participants before the experiment.

Samtykke til deltagelse i studien

Jeg har mottatt informasjonen om studien, og er villig til å delta

____

(Dato, signatur av deltaker)

Appendix D

Information e-mail sent to the volunteers, inviting them to join the project.

Forespørsel om deltakelse i forskningsprosjekt «Statistisk språklæring»

Hei,

Tusen takk for din interesse for mastergradsprosjektet vårt!

Her er en e-post med utfyllende informasjon om prosjektet. Dersom du ønsker å delta ber vi deg om å svare på denne e-posten og oppgi følgende:

- Alder
- Telefonnummer
- Hvilke fremmedspråk du har kjennskap til

Her er mer informasjon om eksperimentet, som vi ber deg lese nøye:

Praktisk info

Eksperimentet vil finne sted i 9. etasje i BB-bygget (Jonas Lies vei 9) ved Haukeland

Universitetssykehus. Deltakelse innebærer ingen fare for fysisk eller psykisk helse. Som deltaker vil du bli bedt om å gjennomføre noen oppgaver på PC, samt fylle ut et kort spørreskjema. Hvert forsøk vil vare i rundt 30 minutter, og du vil få 150,- utbetalt som takk for hjelpen.

Forsøkene vil bli gjort i slutten av oktober/begynnelsen av november 2018. Det vil bli satt av flere uker til forsøkene, og vi vil prøve å være så fleksible som mulig rundt tidspunkt. Vi avtaler tidspunktet nærmere når du melder deg som deltaker.

Bakgrunn for eksperimentet

Dette eksperimentet er en del av et mastergradsprosjekt ved Universitetet i Bergen. Eksperimentet undersøker det som kalles statistisk språklæring, som handler om ubevisst språklæring ved å lytte til et ukjent språk.

Eksklusjonskriterier

For å unngå at uønskede forhold påvirker resultatet i studien har vi et par eksklusjonskriterier.

Kriteriene er hørselsvansker, språkvansker, nevrologiske vansker (som ADHD, oppmerksomhetsvansker eller lignende), andre morsmål enn norsk og kjennskaper til russisk eller andre slaviske språk.

Personer som faller innenfor disse kriteriene blir dessverre ikke inkludert i studien.

Personopplysninger

All innsamlet informasjon om deg som deltar, vil behandles konfidensielt. Innsamlede data vil bli merket anonymt, og er bare tilgjengelig for masterstudentene og veiledere som er involvert i studien.

Når prosjektet avsluttes (etter planen i mai 2019) vil innsamlede personopplysninger bli makulert. Deltakere vil ikke kunne identifiseres ved eventuell publikasjon.

Frivillig deltagelse

Det er helt frivillig å delta i studien, og du kan på hvilket som helst tidspunkt trekke deg uten å oppgi grunn. Alle opplysninger vil bli slettet hvis du trekker deg.

Kontakt:

Prosjektansvarlig:

Professor Arve Asbjørnsen (asbjornsen@uib.no), tlf. 55589084

Masterstudenter:

Andrea Eide (andrea.eide@student.uib.no), tlf. 94798055

Elisabeth Skeistrand (elisabeth.skeistrand@student.uib.no), tlf. 94840073