

KIAP – reflections on a complex corpus

Kjersti Fløttum¹*, Trine Dahl², Anders Alvsåker Didriksen¹ and Anje Müller Gjesdal¹

¹ University of Bergen; ² Norwegian School of Economics

Abstract

In this article we describe the KIAP corpus, a corpus of 450 research articles covering three disciplines (economics, linguistics and medicine) and three languages (English, French and Norwegian) located at the University of Bergen/Uni Research. We describe the rationale for article selection, the functionalities of the corpus as well as important results coming out of the use of the corpus in the KIAP project. Moreover, we discuss the use of the corpus as a methodological tool for the generation and revision of hypotheses on linguistic phenomena characteristic of person manifestation in scientific discourse. Finally, we present some key features of Norwegian medical articles that are relevant in the discussion of the phenomenon of standardisation of scientific discourse.

Keywords: research article; corpus; Norwegian; disciplines; languages

*** Principal contact:**

Kjersti Fløttum, Professor

Department of Foreign Languages, HF-bygget, 5020 Bergen, Norway

Tel.: +47 55 58 22 82

E-mail: kjersti.flottum@if.uib.no

1. Introduction

English academic language has for several decades received much attention from applied linguists and for the past decade or two also from corpus linguists. Descriptions of genres and typical linguistic and discursive features of these genres have been published in dedicated journals (e.g. *Journal of Pragmatics*, *International Journal of Applied Linguistics*, *English for Specific Purposes*, *Journal of English for Academic Purposes*) and books (e.g. Swales 1990; Berkenkotter & Huckin 1995; Hyland 1998, 2000). Some of this research took disciplinary variation into consideration (Hyland 2000), while other studies involved comparisons of English language texts and similar texts in another language (Mauranen 1993; Vassileva 2000). However, few studies did both (for later corpus-based studies doing this, see e.g. Sanderson 2008; Lafuente-Millán, Mur-Dueñas, Lorés-Sanz & Vázquez-Orta 2010). The overarching research question explored in the KIAP project,¹ on the other hand, did relate to both the discipline factor and the language factor. Acknowledging that text is multivoiced, we asked to what extent academic voices, as manifested in research articles, are language-specific or discipline-specific, and how the various voices are realised. The project took its point of departure in features representing traces of SELF (the authors) as well as those of OTHERS (other researchers and readers), in the shape of personal pronouns, metadiscourse, epistemic modifiers, argumentative connectives, bibliographical references, negation and adversatives. In order to examine these questions it was necessary to establish a digital corpus of research articles, reflecting our interest in the interaction of discipline and national languages. The KIAP corpus was established in close collaboration between the AKSIS department at the Unifob research organisation (now Uni computing at Uni Research) and the University of Bergen. In the present paper we would like to reflect on some aspects related to the establishing and use of this complex corpus.

The issues we brought into the project are continuously being discussed in today's research world. The orientation towards standardisation in academic writing becomes more and more pressing, and diversification with disciplinary and national language tradition practices may be increasingly difficult to maintain (Sanderson 2008). The dominant position of the English language in academic discourse is indisputable (Mauranen, Pérez-Llantada & Swales 2010), and the IMRAD structure (**I**ntroduction, **M**ethods, **R**esults and **D**iscussion; see e.g. Swales 1990), originating in experimental science, is close to attaining the same general status in the research article, the most prestigious genre for knowledge dissemination. Has this already led to a loss of specific cultural and disciplinary identities in academic writing practices? To what extent will such a potential loss have an impact for research, higher education and the practices of the students? This situation constitutes a great dilemma for many researchers whose first language is not English and whose writing tradition is not founded on the Anglo-American one. Most of us would like to defend our national and disciplinary traditions, but at the same time we want to publish in internationally renowned outlets promoting standardisation in both language and form.

It was against this background that we started building the KIAP corpus. In this paper we present the project and its main findings (section 2), before we reflect on the corpus as a methodological tool for the generation, development and revision of hypotheses and research questions (section 3). The KIAP project included texts in Norwegian, a small academic language which has received little attention compared to English academic language. We would therefore like to focus specifically on our findings for Norwegian here and contrast these with our findings

¹ For a complete overview of the KIAP (*Kulturell identitet i akademisk prosa/Cultural identity in academic prose*) project, see Fløttum, Dahl and Kinn (2006).

for English and French (section 4). In the final section (5), we attempt to assess the value of relatively small, but specialized corpora like the KIAP corpus, before we outline some perspectives for potential future uses of the corpus.

2. KIAP – project, corpus and general findings

The project “Cultural Identity in Academic Prose”, abbreviated by the Norwegian acronym KIAP, is located at the University of Bergen, but has profited from strong international collaboration. It was financed by the Research Council of Norway during the years 2002–2006. New research is still being undertaken within the project, which (so far) has resulted in 8 master theses and 2 PhD dissertations in Bergen. Through international collaboration, the project is also represented in PhD dissertations outside of Norway (e.g. Poudat 2006; Rinck 2006). This research environment has strongly contributed to the dynamics of the project. Several national and international workshops and symposia have been linked to the project, and in terms of written dissemination of the findings, both the dedicated publication series *Akademisk Prosa: Skrifter fra KIAP* and numerous international publications may be mentioned (some can be found in the References section of the current paper).

Our point of departure for the project was that we considered academic discourse to be clearly rhetorical, a position which we still very much adhere to. Our overall objective has been to give substance to the commonly expressed contestation of the conception of academic discourse as neutral, objective and non-interactive. Several studies have observed the clear presence of the author(s)' voice in academic writings, such as the use of the first person pronoun combined with an argumentative verb – *I/we claim that ...* (e.g. Berge 2003; Breivega 2003; Hyland 1998; Vassileva 2000) In this perspective, we adopted the view that the research article is rhetorical in the sense that it represents a discourse created in order to induce cooperative attitudes and actions (see Prelli 1989), on the one hand, and to give the possibility for the authors to position themselves in or in relation to a particular discourse community, on the other. This position implies the evident fact that the research article is addressed, i.e. directed at someone (Berge 2003). Even if the manifestation of addressivity may be different in different traditions, this observation points to the simple but important fact that the research article constitutes a piece of communication. The primary rhetorical aim of a research article is to create effects which convince the audience to such a degree that the article becomes an integrated part of a particular field's literature, which represents the field's knowledge base.

In order to undertake the relevant analyses, we needed a relatively large corpus. In collaboration with AKSIS (see above), we started the collection of research articles from refereed journals. Since we wanted to pursue a doubly contrastive perspective involving both discipline and language, we collected articles written in three languages, viz. English, French and Norwegian, and within three disciplines, viz. economics, linguistics and medicine. This enabled us to compare, e.g., economics articles written in the three languages as well as, e.g., Norwegian articles within the three disciplines. The corpus design thus made it possible to investigate the importance of both the language factor and the discipline factor, as well as their interaction. The complete electronic corpus consists of 450 articles (about 3,000,000 words) published in refereed journals, in the period 1992–2003. The corpus is divided in 9 subcorpora with 50 articles in each: English economics (*engecon*), English linguistics (*engling*), English medicine (*engmed*); French economics (*frecon*), French linguistics (*frling*), French medicine (*frmed*); Norwegian economics (*noecon*), Norwegian linguistics (*noling*), Norwegian medicine (*nomed*). AKSIS' contribution was crucial in setting up and developing a search function which made it possible to perform searches according to different criteria, such as language, discipline, individual article, article section, etc. The search output also included frequency calculations which became the basis for quantitative analyses (for a detailed description of the KIAP corpus as well as the results of the analyses, see Fløttum et al. 2006).

It is important to emphasise that for the KIAP project, specific linguistic and discursive issues, related to the genre of the research article, formed our point of departure, and also that corpus studies were seen as an important contributor to our *methodological approach*. However, we did not start by developing the corpus and then formulating pertinent research questions which would fit the corpus. We did it the other way around, starting with investigating what would be relevant questions in relation to the purpose of our study. Then, in order to have sufficient data, we proceeded to a systematic collection of research articles, which then were structured in a digital corpus. This corpus thus became an integral part of our method. As regards corpus linguistics, Stig Johansson's definition of the object of corpus linguistics fits well with our own position:

"[...] the object of corpus linguistics is *not* the study of corpora. It is rather the study of language through corpora, [...]." (Johansson 1995:19)

A key issue in the project was to identify possible cultural identities as manifested in linguistic traces of academic voices in the genre of the research article. Within this perspective, our main focus was on person manifestation – what we chose to call the SELF and the OTHER dimensions of academic voices. This focus was specified through three research issues, related to (1) the manifestation of the authors (the SELF dimension) in the texts, (2) the presence of the voices of readers and other researchers (the OTHER dimension) and (3) the presentation of the authors' own research. The focus on person manifestation also explains our choice of theoretical framework, which has been based on a broad interpersonal and polyphonic perspective (see Nølke et al. 2004; Fløttum 2005a), with genre theory as an overarching approach (see for example Berge 2003; Rastier 2001; Swales 1990).

In order to address the three research issues, we selected different linguistic features which may realise the SELF and the OTHER dimensions of person manifestation – in different ways. The main features studied are the following:

- first person and indefinite subject pronouns
- verbs combined with these pronouns
- markers of epistemic modality
- argumentative connectives
- metatextual expressions
- the construction *let us/let me* + infinitive (and its counterpart in French and Norwegian)
- polyphonic constructions (polemic negation and concession)
- bibliographical references

Our main hypothesis throughout the project was that discipline is more important than language in the identification and production of cultural identities; i.e. that there are more similarities between, for example, a Norwegian medical author and a French medical author than between a Norwegian medical author and a Norwegian linguist author. As regards disciplinary differences, a few exploratory analyses undertaken at the beginning of the project

allowed us to draw up three preliminary caricature profiles: the medical researcher as rather absent, hiding behind the text (as in *This study has shown that ...*), the economist as somewhat present but in a modest way (as in *We structure the paper as follows ...*) and the linguist as clearly and polemically present in the text (as in *In contrast to X, we argue that ...*). Our investigations confirmed the general hypothesis in most respects: discipline trumps language. Authors of research articles tend to write more like their disciplinary colleagues writing in other languages than like their language-community co-members writing in other disciplines – with respect to most of the linguistic features studied in the project (see Fløttum et al. 2006).

3. The KIAP corpus as a methodological tool for the development of hypotheses and research questions

In this context, we would like to emphasise the heuristic value of a corpus such as KIAP as a methodological tool. The corpus helped us not only to generate new hypotheses, but also to refocus and adjust our initial hypotheses. The quantitative analyses (concordances, collocations, frequency analyses) as well as the semantic-pragmatic analyses of the corpus revealed several interesting phenomena. In addition, the practical (albeit time-consuming) advantage of having participated in the collection and coding of the corpus material provided us with invaluable insight into the makeup of the texts, sometimes giving rise to unexpected insights. In this section we present an overview of research topics (a–d) that emerged during the corpus analysis (Fløttum 2005b).

*(a) From the general hypothesis of authorial presence manifested by the use of first person pronouns to the hypothesis of authorial presence manifested by the indefinite pronoun **on** in French research articles (see Fløttum 2003, 2004a)*

We still consider first person pronouns to be a relevant and obvious explicit marker of authorial presence in the research article. However, the differences between the use of these pronouns in English and Norwegian on the one hand and in French on the other, turned out to be so important that we soon realised that it was necessary to explore other linguistic constructions in the French research articles that might potentially carry a ‘personal’ value. The analyses showed that the indefinite pronoun *on* (corresponding to English *one*) is highly frequent in the French research articles (for an analysis of indefinite pronouns in English, Norwegian and German, see Johansson 2002; for similar findings on French *on*, see Gjesdal 2003, 2008). Our findings demonstrate a very varied use of this ‘imprecise’ and flexible pronoun. In discourse, the pronoun *on* may potentially correspond to all the personal pronouns in French. Of particular interest in the KIAP context was the use of *on* for author reference, i.e. corresponding to English *I* and Norwegian *jeg/eg* in single-author articles or *we* and *vi* respectively in articles by more than one author, as illustrated in the following example:

(1) Dans un premier temps (paragraphe 1), **on** présentera une liste de caractéristiques, dans l'ensemble bien connues, qui distinguent les deux emplois [...]. (frling06)

(‘Først (i avsnitt 1) vil **vi** presentere en liste over egenskaper, stort sett er velkjente i litteraturen, som skiller de to betydningene [...].’)

(‘First (section 1), we will present a list of characteristics, generally well known, which distinguish the two meanings [...].’)

(b) *From the hypothesis of the absent addressee to the hypothesis of strategies for addressee involvement in structuring and argumentation* (see Fløttum 2003, 2004a; Kinn 2005a, b)

Quantitative searches for linguistic elements explicitly signalling the presence of the addressee, such as second person pronouns, yielded few if any results, as was expected in light of the fact that the research article is not traditionally considered a dialogic genre. However, analyses of various metatextual and metadiscursive expressions pointed to new potential manifestations of the textually elusive addressee. Expressions like “In this article we will first consider” may indicate that this is not just a statement by the author on how the article is structured and what it is about. It may also serve as an invitation to the addressee to participate in developing the arguments and thus in the co-construction of meaning (through an inclusive *we* and a verb referring to an action in which several people can participate). This becomes even clearer with imperative constructions of the type *let us*, as in *let us now go back* in the following example:

(2) Having accounted for basic reconstruction effects with A-movement, let us now go back to the data that motivated Chomsky’s and Lasnik’s conclusion that there are no reconstruction effects with A-movement. (engling49)

These analyses yielded interesting results with regard to how author/addressee interaction may unfold.

(c) *From the hypothesis of explicit presence of other voices than that of the author to the hypothesis of other voices’ implicit presence through polyphonic expressions* (see Fløttum 2005a, 2010, 2012)

The explicit presence of others’ voices in research articles is manifested by direct or indirect reference to results and observations from other researchers’ work, often through some form of reported speech. Our findings in the KIAP project revealed that there are differences between languages and disciplines in terms of how other researchers and their ideas are integrated into a new research text. Through a pilot study early in the project we identified various polyphonic constructions as being particularly interesting in this respect. The voices and viewpoints of other researchers turned out to be implicitly integrated in texts through the use of linguistic features such as polemic *not* and the contrastive and concessive markers *but*. This is a form of subtle interaction that quantitative analyses in turn demonstrated to be a shared and frequent feature of all the nine subcorpora of the KIAP corpus. Through a closer study of the context of the Norwegian syntactic marker of negation *ikke*, we were also led to question the traditional and absolute divide between descriptive and polemical negation. Our analyses indicated that there is a continuum of higher and lower degrees of polemicity, assigning different values to the negation, and determined in part by the genre in question (see Fløttum 2004b).

(d) *From the general hypothesis of authorial presence as manifested by first person pronouns to the hypothesis of author roles manifested by different types of verbs associated with the personal pronoun* (see Fløttum 2004a; Kinn 2005a.)

While items (a)-(c) above represent a further development of our initial hypotheses, item (d) represents a new hypothesis, arising from the concordance-based analyses enabled by the corpus search function. An important objective for the quantitative analyses was to examine our initial hypothesis of first person pronouns as indicative of authorial presence. Quantitative concordance analyses of first person pronouns demonstrated striking regularities in the co-text of this particular feature. Obviously, the verb forms associated with the first person pronouns offer much more information on the nature of author manifestation than the pronoun on its own. An investigation of the verbs involved indicated that at least three *author roles* are enacted in the research article: the author as writer (*I will begin with*) or text guide (*I will return to*), as researcher (*I analyse*) and as 'arguer', i.e. an actor positioning him-/herself (*I claim*). Analyses undertaken by KIAP researcher Torodd Kinn (Kinn 2004, 2005a) revealed an important variation between the subcorpora with regard to the types of verbs that typically co-occur with first person pronouns.

4. Academic discourse in a small language: Norwegian medical articles compared to those of other disciplines and other languages

In this section we will consider how the different features studied in the KIAP project (both SELF and OTHER manifestation) have contributed to the characterisation of Norwegian academic discourse. We in the present context take our point of departure in the discipline of medicine. Medicine represents the most international discipline in our corpus, and the English-language writing tradition in this field is also by far the most standardised, both in terms of structure (IMRAD, see e.g. Swales 1990) and the pervasiveness of the ideal of a non-personal style (see e.g. Gotti & Salager-Meyer 2006).

As already indicated in section 2, the first person pronoun, singular or plural, is the most direct and explicit indicator of author manifestation (SELF dimension). (Since medical research articles are typically multi-author articles, the first person singular becomes irrelevant in this discussion.) However, even if the norm in this discipline clearly favours the use of more impersonal structures, the Norwegian medical subcorpus did comprise some instances of first person plural *vi*. Here is an example of this phenomenon:

(3) *Vi har ikke noen sikker forklaring på denne forskjellen.* (nomed02)

('We do not have any certain explanation of this difference.')

For the corpus as a whole, our quantitative analyses revealed that medical authors generally use fewer first person plural subject pronouns than both economists and linguists (for further details, see Fløttum et al. 2006; Fløttum 2006). However, in terms of the language variable, we found that Norwegian medical authors were similar to their English colleagues in the use of this feature: both groups used more first person plural pronouns than their French peers. In this context, it should be noted that Norwegian authors in general use *vi* (*we*) relatively more frequently than authors writing in English or French.

As mentioned in section 3, indefinite pronouns turned out to play a significant role as indicators of author presence in academic discourse. It is, however, not easy to compare the use of these pronouns in the three languages, since the use of *one* was found to be very limited in the English corpus texts compared to what was observed for the French and Norwegian texts. For French, our results supported the traditional view of the use of *on* as relatively frequent (Gjesdal 2008). For Norwegian, however, the discipline variable turned out to be important, as the use of the indefinite pronoun in medical articles turned out to be quite considerable, and more frequent than in Norwegian economics, but less frequent than in Norwegian linguistics. Here is an example from a Norwegian medical text:

(4) *Man* må unngå manipulasjon av hjertet før [...]. (nomed15)

(‘One must avoid manipulation of the heart before ...’)

Another feature serving as a marker of author manifestation is metatext. In the KIAP project we searched for various expressions referring to (parts of) the text itself. By expressions such as *in this article* the author guides the readers through the text (see Dahl 2003, 2004) and explains what will be done and where (for example *In section 4 we analyse ...*). For methodological reasons, the automated searches and the quantitative analyses were limited to the following expressions (and their corresponding expressions in French and Norwegian): *article, paper, (sub)section, above, now, below*. Our results indicated very clearly that in medical articles this feature is more or less absent in all the three languages. In our opinion, this finding may be explained by reference to the highly standardised text structure in this discipline, making metatext superfluous. For comparative purposes, we may add that articles in Norwegian economics contain a relatively high number of metatextual expressions.

One reason for integrating polyphonic theory in the analysis of the KIAP corpus was the possibility this perspective offers to bring in both the SELF and the OTHER perspective. An obvious example of this is polemic negation where the author implicitly refutes another point of view than his or her own. The negation expressions studied in our project were *not* in English, *ne ... pas* in French and *ikke/ikkje* in Norwegian, all indicating polyphony through the refutation of an underlying positive point of view. The overall quantitative results showed, roughly, that linguists use more negation than economists, and economists use it more than medical researchers. As regards our focus discipline in this section, we note that the language factor is also important, in that Norwegian and English medical authors use negation more often than French medical authors. Here is a Norwegian example:

(5) [...], talar alle desse tre forholda for at den observerte auken *ikkje* kan tilskrivast mediaoppstyret og påfølgjande seponering av pillen Marvelon. (nomed10)

(‘[...], all these conditions indicate that the observed increase can not be related to [...]’)

It may also be noted that Norwegian authors in general use negation more often than authors writing in English and French. This led us to ask whether authors writing in Norwegian might be more polemic than their colleagues writing in other languages.

The use of connectives may also be interesting in the perspective of author manifestation. However, the comparison of connectives in different languages requires particular semantic-pragmatic

considerations. Even if connectives correspond more or less to each other across languages, they often have particular semantic-pragmatic values in different languages (see e.g. Didriksen 2010). We limited the automated searches to the adversative connective *but* and its corresponding connectives in French (*mais*) and Norwegian (*men*).² These may also convey different meanings; however, they all have a concessive meaning in common, and this was the meaning which was particularly interesting in relation to our research questions. In their concessive capacity, these connectives function as markers of polyphony as in the following example:

(6) Flere arbeider har vist at dette gir bedret prognose, *men* medikamentet er kostbart og gir økte blødningskomplikasjoner. (nomed01)

(‘Several investigations have shown that this gives a better prognosis, but the medication is expensive and causes complications [...].’)

The author here first makes a concession related to a proposition expressed in the sequence preceding *but* (the source of this proposition is not necessarily the author). The author then contrasts the proposition with a succeeding sequence, presented as the most important (the point of view of the author here and now). In this way, concessive constructions contribute to author manifestation and even author positioning.

The quantitative data related to the frequency of the connectives in question revealed that linguistics is the discipline that displays the highest frequency of concessive constructions, while medicine is the discipline with the lowest relative frequency. However, Norwegian medical texts displayed a remarkably high frequency of this feature compared to English and French texts in this discipline.

Let us now look at the presence of bibliographical references, which are the most obvious markers of the OTHER dimension. We limit our presentation of this feature here to the main figures related to the frequency of bibliographical references, without taking the different types of references and their function in the articles into consideration (see Fløttum & Rastier 2003). Without going into detail about the importance of bibliographical references in all kinds of academic discourse, we just note that medicine is a discipline with a relatively uniform referencing standard (Gotti & Salager-Meyer 2006). When analysing the frequencies of bibliographical references (whether introduced by a number or by author name + year of publication), we find that when considering the three languages together, medical authors use more references than economists and linguists. However, considering the languages separately, it turns out that Norwegian medical authors use more references than their English and French colleagues.

To sum up some of the findings for the KIAP project as a whole: We found that regarding the issue of cultural identity, discipline has greater influence than language. This means that, for example, there are more similarities between Norwegian and French medical articles than between Norwegian medical and linguistics articles.³ Our quantitative analyses revealed that both discipline and language have an effect on the frequency of all the main phenomena studied in the project. For most of them, however, discipline seems to be more important than language (see Table 1).

² For quantitative findings on a range of French connectives in the French part of the KIAP corpus, see Didriksen (2004).

³ For more on language and cultural differences, see Fløttum 2012.

Table 1

Effects of the discipline and language factors and their interaction on variance in the KIAP corpus

Feature	Discipline	Language	Interaction	Residual
Metatex	52%	12%	2%	34%
Bibliographical references	44%	5%	2%	49%
Adversative conjunctions	25%	9%	1%	65%
Negation	18%	18%	0%	64%
First person subjects	15%	11%	1%	73%
Indefinite pronoun subjects	9%	49%	3%	39%

Another notable general finding is that we have observed clear individual differences between articles even within one discipline and within one language, indicating that the genre of the research article to some extent allows academic authors to develop their own ‘voice’ when writing up their research results. In a study related to the KIAP-project (Didriksen & Gjesdal 2006), the relationship between individual linguistic variation and genre conventions was examined through a quantitative, corpus-based study of the use of the French first person singular pronoun *je* in research articles. This study shows that although there is considerable individual variation, authors still tend to comply with the norms of the genre. Results also indicate that individual language practices vary over time.

As regards medical discourse specifically, our findings indicate, in very general terms, that it is clearly different from economics and linguistics discourse. The most obvious difference is related to author presence. Medical authors are not very visible in their texts compared to their economist and linguist peers. They ‘hide’ behind passive constructions and impersonal formulations of different kinds. Scientific observations ‘present themselves’ to a larger extent than in economics and linguistics articles. In an overall perspective, we may also conclude that medical articles resemble each other to a certain extent, whatever language they are written in. However, as already indicated, we have seen that Norwegian medical authors to some extent exploit the studied features of SELF and OTHER representation somewhat differently from their colleagues in the other two writing cultures. They are more clearly and directly present than their English and especially French colleagues as regards the use of the first person pronoun, the indefinite pronoun, negation and concessive *but*. Nomed is also the subcorpus which displays the highest relative frequency of bibliographical references.

At this point it seems fitting to include some reflections on the nature of our various language subcorpora. While the English texts in all three disciplines are aimed at an international audience, the French texts attract a somewhat smaller readership. As for the Norwegian texts, their only potential audiences outside of Norway are found in the other Scandinavian countries. Our basis for selecting the Norwegian texts was also smaller than for French and much smaller than for English. In fact, studies from the early 2000s have shown that eight out of ten articles written by Norwegian researchers are written in English and more than a third of these were published outside of Norway (Schwach 2004, see also De Smedt et al. 2012). In consequence, the available material was relatively limited for Norwegian. Only one journal was available for the disciplines of medicine and economics, while there were two for linguistics. Notably in economics and medicine, the Norwegian articles tended to address a broader and more diverse audience than just the relevant research community. Practitioners in both

fields (e.g. medical doctors outside research hospitals and economists working in government departments) also published in these journals and were also part of the intended readership. This fact may also have influenced the findings for the Norwegian subcorpora for these two disciplines.

5. Final remarks

While KIAP may be denoted as a large and representative corpus of scientific discourse, the rapid developments in corpus linguistics and the advent of massive data bases and collections imply that the landscape of corpus studies is changing. However, we believe that carefully composed corpora developed from philological considerations still offer benefits that massive collections of raw data do not (Rastier 2011). An important value of the KIAP corpus, which (despite being considered large in terms of specialised corpora) more generally can be said to be of limited size, is that it is possible to get a relatively clear picture of the corpus as a whole. It is in fact possible to access each text, get an overview, and relate observations made there to the entire corpus. There are sometimes objections to the use of electronic corpora, for instance that “you only find what you search for” (Fløttum 2005b). However, in the KIAP project, we repeatedly experienced that new hypotheses and questions developed through the work on the corpus (see above). Furthermore, the possibility to study each individual article in depth justifies a rejection of such objections.

Finally, we would like to return to the issues we raised at the beginning of the article, i.e. the status of scientific discourse in smaller language communities in the face of increasing standardization and pressure to publish in English. The KIAP corpus provides a window into scientific discourse of the late 90s and early 00s, and is therefore well suited as a basis for diachronic analyses and comparison with academic discourse today, ten years later. It would be interesting to compare the analyses of the KIAP corpus with those of a similar corpus of today’s academic discourse, in order to observe which changes – if any – the genre of the research article has undergone in this period. Has the trend towards standardisation continued, or has increasing attention to the potential problems associated with standardisation influenced writing practices? Furthermore, the resource availability for Norwegian corpus linguistics has also increased significantly over this ten year period, and large corpora are now available through the Norwegian National Library’s digital language resource collection, *Språkbanken*. Thus, it is now possible to use resources like the Norwegian Newspaper corpus (Andersen & Hofland 2012) to compare the KIAP corpus of academic discourse with corpora of general language.

References

- Author, A. A. (Year of publication). *Title of work: Capital letter also for subtitle*. Location: Publisher.
- Berndt, T. J. (2002). Friendship quality and social development. *Current Directions in Psychological Science*, 11, 7-10.
- Andersen, G., Hofland, K. (2012). Building a large corpus based on newspapers on the web. In: G. Andersen (ed), *Exploring Newspaper Language. Using the web to create and investigate a large corpus of modern Norwegian*. Amsterdam: John Benjamins, 1-28.
- Berge, K. L. (2003). The scientific text genres as social actions: Text theoretical reflections on the relations between context and text in scientific writing. In: K. Fløttum, F. Rastier (eds), *Academic discourse. Multidisciplinary approaches*. Oslo: Novus, 141-157.
- Berkenkotter, C., Huckin, T. N. (1995). *Genre Knowledge in disciplinary communication*. Hillsdale, N.J.: Lawrence Erlbaum.
- Breivega, K. R. (2003). *Vitskaplege argumentasjonsstrategiar. Ein komparativ analyse av superstrukturelle konfigurasjonar i medisinske, historiske og språkvitskaplege artiklar. Sakprosa 8*. Oslo: Norsk sakprosa.
- Dahl, T. (2003). Metadiscourse in research articles. In: K. Fløttum, F. Rastier (eds), *Academic discourse. Multidisciplinary approaches*. Oslo: Novus, 120-138.
- Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36, 1807-1825.
- De Smedt, K., Lyse, G. I., Gjesdal, A. M. & G. S. Losnegaard (2012). *Norsk i den digitale tidsalderen*. Düsseldorf: Springer.
- Didriksen, A. A. (2004). *Donc* et autres connecteurs argumentatifs dans les articles de recherche français. Master thesis. Department of Romance Studies, University of Bergen.
- Didriksen, A. A. (2010). La méthode des miroirs sémantiques : un point de départ pour l'identification des instructions logico-sémantiques d'un connecteur. In: J. Havu (ed), *Actes du XVIIe Congrès des romanistes scandinaves - Actas del XVII Congreso de romanistas escandinavos*. Tampere University Press, 238-252.
- Didriksen, A. A., Gjesdal, A. M. (2006). Genre Constraints and Individual Linguistic Variation. In: E. Suomela-Salmi, F. Dervin, F. (eds) *Cross-cultural and Cross-linguistic Perspectives on Academic Discourse*. Turku: Department of French Studies, The University of Turku, Finland, 47-57.
- Fløttum, K. (2003). Personal English, indefinite French and plural Norwegian scientific authors? Pronominal author manifestation in research articles. *Norsk Lingvistisk Tidsskrift* 21, 21-55.
- Fløttum, K. (2004a). La présence de l'auteur dans les articles scientifiques: étude des pronoms *je, nous* et *on*. In: A. Auchlin et al. (eds), *Structures et discours*. Québec: Ed. Nota Bene, 401-416.
- Fløttum, K. (2004b). Polyfonisk interaksjon via IKKE i vitenskapelig diskurs. *Rhetorica Scandinavica* 31, 23-40.
- Fløttum, K. (2005a). The self and the others – polyphonic visibility in research articles. *International Journal of Applied Linguistics* 15, 29-44.

- Fløttum, K. (2005b). Lingvistiske og diskursive studier i KIAP-korpuset. Om utvikling og bruk av et flerspråklig og flerdisiplinært korpus av vitenskapelige artikler. In: S. Rike (ed), *Korpus i forskning og undervisning. Hvor står vi i Norge?* Kristiansand: Høgskolen i Agder, 25-51.
- Fløttum, K. (2006). En språklig analyse av norsk medisinsk prosa. *Tidsskrift for Den norske lægeforening*, 126 (1), 65-68.
- Fløttum, K. (2010). Linguistically marked cultural identity in research articles. In: G. Garzone, J. Archibald (eds), *Discourse, identities and roles in specialized communication*. Bern: Peter Lang, 267-280.
- Fløttum, K. (2012). Variation of stance and voice across cultures. In: K. Hyland, C. Sancho Guinda (eds), *Stance and Voice in academic discourse*. London: Palgrave MacMillan, 218-231.
- Fløttum, K., Rastier, F. (eds). (2003). *Academic discourse. Multidisciplinary approaches*, Oslo: Novus.
- Fløttum, K., Dahl, T., Kinn, T. (2006). *Academic Voices*. Amsterdam/Philadelphia: John Benjamins.
- Gjesdal, A. M. (2003). L'emploi du pronom "on" dans les articles de recherche. Une étude diachronique et qualitative. Master thesis. University of Bergen: Romansk institutt.
- Gjesdal, A. M. (2008). *Étude sémantique du pronom ON dans une perspective textuelle et contextuelle*. PhD dissertation. Bergen: University of Bergen.
- Gotti, M., Salager-Meyer, F. (eds). (2006). *Advances in Medical Discourse Analysis: Oral and Written Contexts*. Bern/Berlin: Peter Lang.
- Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.
- Hyland, K. (2000). *Disciplinary discourses: social interactions in academic writing*. Harlow, UK: Longman.
- Johansson, S. (1995). *Mens sana in corpore sano: On the Role of Corpora in Linguistic Research*. *The European English Messenger*, IV, 2, 19-25.
- Johansson, S. (2002). Viewing language through multilingual corpora, with special reference to the generic person in English, German and Norwegian. In: L.I. Rábade, S.M. Doval Suarez (eds), *Studies in Contrastive Linguistics*. Universidade de Santiago de Compostela, 515-554.
- Kinn, T. (2004). Cognitive research agents in academic prose. *Akademisk Prosa: Skrifter fra KIAP*. Romansk Institutt, Universitetet i Bergen, 137-149.
- Kinn, T. (2005a). "Denne artikkelen analyserer ...". Den tenkjende forskeren i norske forskingsartiklar. *MONS 10 utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003*. Kristiansand: Høgskoleforlaget, 183-194.
- Kinn, T. (2005b). Tilbod og innbydingar: imperativ med *la* i forskingsartiklar. *Norsk lingvistisk tidsskrift* 23 (2), 145-175.
- Lafuente-Millán, E. Mur-Dueñas, P. Lorés-Sanz, R., Vázquez-Orta, I. (2010). Interpersonality in written academic discourse: Three analytical perspectives. In: R. Lorés-Sanz, P. Mur-Dueñas, E. Lafuente-Millán (eds), *Constructing interpersonality: Multiple perspectives on written academic genres*. Newcastle: Scholars Publishing, 13-39.
- Mauranen, A. (1993). *Cultural differences in academic rhetoric*. Frankfurt am Main: Peter Lang.

- Mauranen, A., Pérez-Llantada C., Swales, J. (2010). Academic Englishes – A Standardized knowledge? In: A. Kirkpatrick (ed), *The Routledge handbook of world Englishes*. London: Routledge, 634-652.
- Nølke, H., Fløttum, K., Norén, C. (2004). *ScaPoLine. La théorie scandinave de la polyphonie linguistique*. Paris: Kimé.
- Poudat, C. (2006). *Étude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres*. Doctoral dissertation. Orléans: Université d'Orléans.
- Prelli, L. J. (1989). *A Rhetoric of science: Inventing scientific discourse*. University of South Carolina Press.
- Rastier, F. (2011). *La mesure et le grain : sémantique de corpus*. Paris: Honoré Champion.
- Rastier, F. (2001). *Arts et sciences du texte*. Paris: PUF.
- Rinck, F. (2006). *L'article de recherche en Sciences du langage et en Lettres. Figure de l'auteur et identité disciplinaire du genre*. Doctoral dissertation Grenoble: Université de Grenoble.
- Sanderson, T. (2008). *Corpus, culture, discourse*. Tübingen: Gunter Narr Verlag.
- Schwach, V. (2004). *Norsk vitenskap – på språklig bortebane? Et pilotprosjekt om språkbruk blant fagsamfunnet av forskere i Norge*. NIFU skriftserie, 9. [<http://www.nifu.no/Norway/Publications/2004/skriftserie9-2004.pdf>, [accessed 10 September 2012].
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Vassileva, I. (2000). *Who is the author? A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian academic discourse*. Sankt Augustin: Asgard Verlag.