

DRAFT

Too chatty: Learner academic writing and register variation

Authors : Gaëtanelle Gilquin & Magali Paquot

Affiliation : Université catholique de Louvain

Running head: Too chatty: Learner academic writing and register variation

Contact details :

Gaëtanelle Gilquin
Université catholique de Louvain
Centre for English Corpus Linguistics
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-Neuve
Belgium
gaetanelle.gilquin@uclouvain.be
tel. +32 10 47 92 59
fax +32 10 47 40 34

Magali Paquot
Université catholique de Louvain
Centre for English Corpus Linguistics
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-Neuve
Belgium
magali.paquot@uclouvain.be
tel. +32 10 47 40 34
fax +32 10 47 40 34

Abstract

The study reported on in this paper uses corpus data in order to examine how upper-intermediate to advanced EFL learners from a wide range of mother tongue backgrounds perform a number of rhetorical functions particularly prominent in academic discourse, and how this compares with native academic writing. In particular, it is shown that one of the problems experienced by EFL learners is that they tend to use features that are more typical of speech than of academic prose, which suggests that they are largely unaware of register differences. Four possible explanations are offered to account for this register confusion, namely the influence of speech, L1 transfer, teaching-induced factors and developmental factors.

1. Introduction

Flowerdew (2001) highlights three areas of difficulty in learner academic writing, viz. collocational patterning, pragmatic appropriacy and discourse features. Collocational patterning has to do with learners' lack of familiarity with the typical lexico-grammatical environment of words (e.g. **We have **performed** a survey*, **A questionnaire has been **conveyed** to the public*). Pragmatic appropriacy concerns the writer's attitude to his/her message, as realised through modal verbs, modal adjuncts, boosters, hedges, etc. Hyland & Milton (1997), for example, have shown that Cantonese learners have considerable difficulty conveying an appropriate degree of qualification in their academic essays. In this paper, we will focus on the third area of difficulty, namely discourse features, and more specifically on "the 'discourse value' of lexical and structural items in context" (Flowerdew 2001: 374) – although pragmatic inappropriacy will also emerge from the analysis, as a consequence of the overuse, underuse or misuse of some discourse features. In particular, we will investigate the extent to which upper-intermediate to advanced learners of English use spoken-like features in their academic prose. While this tendency has already been brought to light for specific words and for learners from a number of mother tongue (L1) backgrounds, as will be briefly shown in Section 2, we will demonstrate in Section 3 that this is actually a more general problem, concerning learners from many L1 populations and pertaining to many of the rhetorical functions that serve to organise academic discourse. In Section 4, we will provide

four possible explanations for this lack of register-awareness, namely the influence of speech, L1 transfer, teaching-induced factors and developmental factors.

2. Learners' confusion of registers

While language has long been seen as a basically unitary phenomenon, corpus-based studies such as Biber's (1988) have brought to light the great variation that exists within a language. Different genres have different characteristics, and speech cannot be fully described by a grammar of the written language (see e.g. Carter & McCarthy 1995 for English). Producing a text in a given genre therefore requires what Lorenz (1999: 64) calls "text-type sensitivity". As underlined by Lorenz and others, however, non-native writers often lack such sensitivity and may struggle to find an appropriate voice, that is, a style that is suitable for a particular text type.

In particular, it has been repeatedly noted that learners often do not respect the differences between speech and writing, and that the devices they use "can give confusing signals of register" (Field & Yip 1992: 26). Thus, using a fully automatic method of investigation, Granger & Rayson (1998) have shown that French-speaking learners overuse many lexical and grammatical features typical of speech, such as first and second person pronouns or short Germanic adverbs (*also, only, so, very, etc.*), and significantly underuse many of the characteristics of formal writing, such as a high density of nouns and prepositions. Other studies, focusing on more specific items, have brought to light the same problem. For example, Lorenz (1999) discusses the marked overuse of the conjunction *because* and the adverb *so* in German learner writing; French, Spanish and Swedish learners' heavy reliance on *I think* to express their personal opinion is reported by Granger (1998), Neff et al. (2007) and Aijmer (2002) respectively; foreign learners' overuse of *of course* is highlighted by Narita & Sugiura (2006) for Japanese, Granger & Tyson (1996) for French and Altenberg & Tapper (1998) for Swedish.

In contrast to the above-mentioned studies, which focus on learners from one or two L1 populations, the current study examines learners from a large number of mother tongue backgrounds. It takes as a starting point the expression of rhetorical functions in academic writing, as explained in the next section.

3. Spoken-like features in academic discourse functions

When writing an academic essay, one has to fulfil a number of rhetorical functions in order to organise one's discourse. These functions include: expressing possibility and certainty, introducing new topics, establishing causal links, quoting other people's work and concluding a piece of writing. Within the framework of a close collaboration between the Centre for English Corpus Linguistics (Université catholique de Louvain) and Macmillan Education on the second edition of the *Macmillan English Dictionary for Advanced Learners*, we studied twelve of these functions and compared the way they are performed in native and non-native academic English (Gilquin et al. 2007; see also Gilquin et al. in press for a detailed description of this project). In this section, we will focus on what this comparison tells us in terms of register, and more particularly, how learners' performance can be situated along the speech-writing continuum.

3.1. Data and methodology

Our study started with the collection of a number of words and phrases used in rhetorical functions particularly prominent in academic writing. A large proportion of the lexical items examined come from Paquot's (2007) *Academic Keyword List*. Paquot applied a number of criteria including keyness, range and evenness of distribution (cf. Rayson 2003) on three corpora of academic texts from several disciplines (e.g. arts, social science, applied science, technology and engineering) to select EAP-specific words and phrases, i.e. a rather formal vocabulary common to a wide range of academic texts from different disciplines. The *Academic Keyword List* includes 930 lemmas, examples of which are given in Table 1. From this list, we selected words and phrases which are used to perform one of the following functions: adding information, comparing and contrasting, exemplifying, expressing cause and effect, expressing personal opinion, introducing a concession, introducing topics and related ideas, listing items, reformulating, quoting and reporting, summarizing and drawing conclusions. We added certain items which did not emerge from Paquot's (2007) corpus-driven analysis but are commonly mentioned in academic textbooks as serving specific rhetorical functions (e.g. *in brief*, *in a nutshell* or *all in all* for summarizing and drawing conclusions). The final list of words and phrases for this study included a total of about 350 items.

Table 1 about here

The use of these lexical items was then analysed and compared in three different corpora, namely a corpus of academic native English, a corpus of spoken native English and a corpus of learner English. The native data come from the academic and spoken components of the *British National Corpus* (BNC), totalling 15 million and 10 million words, respectively. The academic corpus (BNC-W) consists of samples from books and journal articles in several disciplines, e.g. humanities and arts, medicine, natural science, politics, law and education. The spoken corpus (BNC-S) includes a wide variety of spoken registers, among which broadcast documentary and news, interviews and lectures. The learner data come from the second version of ICLE, the *International Corpus of Learner English* (Granger et al., forthcoming), which includes 3.5 million words of academic writing – argumentative and literary essays – produced by upper-intermediate to advanced¹ foreign learners from 16 mother tongue backgrounds belonging to different language families (e.g. French, Chinese, Norwegian, Turkish). ICLE is encoded with many learner and task variables (see Granger 2003). Because settings are extremely important in learner writing, as convincingly demonstrated by Ädel (forthcoming), we limited our analysis to argumentative essays written in untimed conditions and with access to reference tools. These criteria reduced our sample to some 1.5 million words and 14 L1 subcorpora (two of the components, the Bulgarian and Tswana components, did not contain enough data meeting the criteria and were therefore left out). The 14 subcorpora are the following: Chinese (ICLE-CH), Czech (ICLE-CZ), Dutch (ICLE-DU), Finnish (ICLE-FI), French (ICLE-FR), German (ICLE-GE), Italian (ICLE-IT), Japanese (ICLE-JP), Norwegian (ICLE-NO), Polish (ICLE-PO), Russian (ICLE-RU), Spanish (ICLE-SP), Swedish (ICLE-SW) and Turkish (ICLE-TU).

Researchers such as Lorenz (1999) and Hyland & Milton (1997) criticise the use of professional writing in learner corpus research, arguing that it is “both unfair and descriptively inadequate” (Lorenz 1999: 14) and taking a stand against the “unrealistic standard of ‘expert writer’ models” (Hyland & Milton 1997: 184). Native student writing is arguably a better type of comparable data to EFL learner writing if the objective of the comparison is to describe and evaluate interlanguage(s) as fairly as possible. It is, however, highly questionable whether findings from such comparisons can make their way into the

¹ Learners’ level in ICLE was determined on the basis of external criteria (the students were in their third or fourth year of university study). Assessing some 200 texts from ICLE on the basis of internal criteria revealed proficiency levels ranging from B2 to C2 in the Common European Framework (but with a majority of C1), cf. Granger & Thewissen (2005).

classroom. As Leech (1998: xix) puts it, “[n]ative-speaking students do not necessarily provide models that everyone would want to imitate”. For example, native students have been shown to produce more dangling participles than EFL learners (Granger 1997) and different types of orthographic errors (Cutting 2000). The question of the norm can be settled by taking into account the aim of the comparison. Professional writing has a major role to play in learner corpus research as soon as pedagogical applications are considered. As noted by Ädel (2006: 206-207):

On the one hand, it can be argued that in order to evaluate foreign learner writing by students justly, we need to use native-speaker writing that is also produced by students for comparison. On the other hand, it can also be argued that professional writing represents the norm that advanced foreign learner writers try to reach and their teachers try to promote. In this respect, a useful corpus for comparison is one which offers a collection of what Bazerman (1994: 131) calls ‘expert performances’.

We nevertheless agree with De Cock’s (2003: 196) comment that “argumentative essay writing has no exact equivalent in professional writing”. Special care will thus be taken to interpret results in the light of genre analysis, as differences between student essays and expert writing may simply reflect differences in their communicative goals and settings (cf. Neff et al. 2004).

The frequencies of the selected words and phrases in the 14 subcorpora of ICLE were compared to those of the same items in the written and spoken native corpora, which made it possible to highlight cases of overuse (when learners use more of a particular item than native speakers) and underuse (when they use less of a particular item than native speakers).² Our objective was to identify cases of overuse and underuse that are shared by learners from a wide range of mother tongue backgrounds and which can be used to inform generic pedagogical tools. In what follows, we report on some of the register-related problems that were found to characterise a majority of the learner corpora.

3.2. Some results

The analysis of the learner data and their comparison with the written and spoken native corpora reveal a strong tendency among learners, regardless of their mother tongue, to use

² It is important to note that the terms “overuse” and “underuse” are descriptive, not prescriptive, terms: they merely refer to the fact that a linguistic form is found significantly more or less in the learner corpus than in the reference corpus. Chi square tests were performed in order to establish the statistical significance of these under- and overuses.

spoken-like features in their written production. They tend to overuse words and phrases which are more likely to appear in speech, and underuse more formal expressions typical of academic writing. In addition, they sometimes use conjunctions in positions favoured by spoken registers, rather than academic prose.

A first example of EFL learners' tendency to use spoken-like items in their academic essays is found in the way learners express possibility, and more particularly, in the distribution of the adverbs *perhaps* and *maybe*. As appears from Figure 1, *perhaps* is much more common in native academic writing than *maybe* (347 occurrences of *perhaps* per million words [pmw] vs. 13 *maybe* pmw). A comparison of academic texts reveals that, although frequencies may differ significantly across academic disciplines, *perhaps* is always more frequent than *maybe* (e.g. in humanities: 550 *perhaps* pmw vs. 19.6 *maybe* pmw; in medicine: 160.3 *perhaps* pmw vs. 2.8 *maybe* pmw; in natural science: 218.6 *perhaps* pmw vs. 5.4 *maybe* pmw). Learner writing, however, exhibits just the opposite tendency, with a higher relative frequency of *maybe* than of *perhaps* (342 *maybe* pmw vs. 285 *perhaps* pmw). Since, as the graph shows, *maybe* is more prominent in speech than in writing, learners may be said to favour an adverb that is more typical of speech than of writing, as illustrated by (1). *Perhaps*, by comparison, is less frequently used, although it would be better suited to express possibility in academic writing, as would a whole range of lexical items, also underused by the learners, including *apparently*, *it seems that* or *it is possible that*.

Figure 1 about here

- (1) In general, **maybe** English really is slightly easier to learn than other languages but it is not as simple that one can learn it overnight³. (ICLE-PO)

More generally, the corpus analysis reveals that learners usually underuse items that express possibility (e.g. *apparently*, *presumably*, *likely*, *assumption*), but overuse items expressing certainty. This is particularly clear in the case of adverbs, with adverbs such as *really*, *of course*, *certainly*, *absolutely* or *definitely* figuring much more prominently in learner writing than in native writing, as schematised in Figure 2 and exemplified in (2) and (3). Again, as Figure 2 shows, these amplifying adverbs are typical of speech. In academic writing, by contrast, they are less frequently used, which can be related to the fact that

³ Learners' sentences are reprinted as they appear in ICLE, with no corrections made to their writing.

“[a]cademic writing is typically cautious: sweeping generalizations are avoided and instead claims are limited or qualified” (Blanpain 2006: 45). Tentativeness, with the use of so-called hedges, has a key role in academic writing. By making room for other positions than the one advocated by the writer, it helps create good relationships between the writer and the reader, while at the same time allowing the writer to avoid “personal accountability for statements” (Hyland 1994: 240). For these reasons, “[l]earning to use hedging devices appropriately, without becoming over-tentative, is an essential academic skill” (Blanpain 2006: 45).

Figure 2 about here

- (2) Do you think father of the baby can tolerate for a long time for the cries of the baby. I think the answer is clear; **of course** not, is not it? But mother can tolerate **of course**. Mothers have deeper feeling of mercy and they will **absolutely** try the baby to stop crying. (ICLE-TU)
- (3) So unfortunately money **really** takes a very important, supreme part in our life. And we **certainly** can not talk about freedom in the world until we understand how much evil depends on this role of money. (ICLE-RU)

Another issue which learners seem to have problems with is the extent to which they should be visible as writers. As noted by Recski (2004), in the case of argumentative essays such as those contained in ICLE, “personal references and subjective attitudes are certainly hard to avoid”, since learners are explicitly prompted to give their personal opinions (topics for the essays include: “Some people say that in our modern world, dominated by science, technology and industrialism, there is no longer place for dreaming and imagination. What is your opinion?” and “In the 19th century, Victor Hugo said: ‘How sad it is to think that nature is calling out but humanity refuses to pay heed.’ Do you think it is still true nowadays?”). It should be borne in mind, however, that personal opinions are present in academic prose too. By using impersonal structures such as *it is reasonable to* and *it is worth noting that*, or adverbs such as *interestingly* or *surprisingly*, professional writers express their attitude towards their message. Learners, by contrast, tend to make their presence strongly felt in their writing. They often express personal opinions with the phrase *it seems to me*, as in (4) and (5), where native speakers would probably use the simple expression *it seems* instead, cf. (6). Similarly, they tend to use explicit expressions such as *I think that* or *to my mind*, which are in

many cases unnecessary. Even if these expressions were not used in sentences (7) and (8), it would still be clear that the writer is expressing his/her own opinion.

- (4) **It seems to me** that being tolerant requires strength of character and also a certain maturity. (ICLE-GE)
- (5) **It seems to me** a paradox, I mean what high commands expect and what they sometimes obtain. (ICLE-SP)
- (6) **It seems** that these authors see mind as an entity distinct from the processes that support it. (BNC-W)
- (7) In conclusion, **I think that** capital punishment is not the most effective way to punish murderers. (ICLE-CH)
- (8) Others exaggerate dieting in order to keep their slender figure. **To my mind** it is stupid to slim down until you are emaciated. (ICLE-GE)

Learners are also often too visible when introducing new topics or ideas. They frequently use the expression *I would like / want / am going to talk about...*, as illustrated in (9). Native writers, on the other hand, prefer more impersonal devices, e.g. *this article examines...* or *topics addressed will include...*

- (9) The next effect **I would like to talk about** is social contact and it is best illustrated by these words: “before I bought my dog I did not know my neighbours, but now everybody greets me and my dog.” (ICLE-SW)

This high writer visibility is more characteristic of speech than of writing. The expressions *it seems to me* and *I would like / want / am going to talk about...*, for example, are very frequent in speech, but much less so in academic writing, as Figures 3 and 4 show. This, again, may have unfortunate consequences for learner essays peppered with such expressions. As further noted by Recski (2004), because signals of the writer’s presence are typically used in speech to “illustrate the involvement of the speaker in what she or he is saying”, their use in writing “provoke[s] similar feelings in the readers (...). The readers feel that the writer is somehow involved in the text and automatically attribute the writer of the (otherwise impersonal) text with personal, subjective emotions and attitudes – as they would attribute them to a speaker they could see or hear” (ibid.). By avoiding such signals, one can help develop the “sense of impersonal objectivity” (ibid.) that is expected of academic prose.

Figure 3 about here

Figure 4 about here

Learners' tendency to favour more spoken-like features may also concern the positioning of words. While Conrad (1999) suggests that the unmarked position for linking adverbials in both conversation and academic prose is at the beginning of a sentence, some items show different preferences depending on register. Thus, *and* is rarely used sentence-initially in academic prose, but this position is very common in speech, as Figure 5 makes clear. Learners, by overusing the sentence-initial positioning, cf. (10), therefore give an oral tone to their essays.

Figure 5 about here

- (10) Some soldiers were sent there even without knowing, where they were going.
And some soldiers even couldn't shoot well.

Interestingly, most of the rhetorical functions we examined turn out to be characterised in learners' essays by a number of items which are more typical of speech than of writing (see Paquot 2007). Table 2 summarises these findings, and sentences (11) to (13) give some examples from ICLE. In the next section, we will try to understand why learner writing exhibits so many of these spoken-like features.

Table 2 about here

- (11) For instance when I watch movies **like** "The Godfather" or "The Untouchables" I often drift into dreams of being either Michael Corleone (The Godfather) or Al Capone (The Untouchables). (ICLE-NO)
- (12) We should know that the person whom we get acquaintance by E-mails is not always the person we expect to be. **So**, we should not connect the situation in the virtual world with reality too easily. (ICLE-JP)
- (13) These events seldom or never happen in reality, **though**. (ICLE-BU)

4. Explaining the spoken-like nature of learner writing

In what follows, we will put forward four possible explanations to account for the spoken-like nature of learner writing. We will start with what may seem to be the most straightforward explanation, the influence of the spoken medium, and will argue that it is not a very likely explanation in this case. We will then present three other explanations – which are not mutually exclusive – namely the influence of the mother tongue, the influence of teaching and the effect of developmental factors.

4.1. The influence of speech

Learners of English as a Second Language (ESL) have been shown to borrow oral strategies in their writing (see Schleppegrell 1996 or Hinkel 2003). One reason for this seems to be that those students “have learned English primarily through oral interaction, rather than through instruction” (Schleppegrell 1996: 274). When attempting to write academic discourse, they resort to what they know best, namely speech, which results in features inappropriate for academic writing.

The situation is quite different for learners of English as a Foreign Language (EFL), who mainly learn English through instruction. For them, spoken input tends to be restricted to teacher talk and oral teaching materials. Teacher talk may be relatively limited, depending on the teaching method (use of the mother tongue or not, focus on form or on communicative competence). In addition, for many learners teacher talk is non-native teacher talk, and a non-native teacher is at best a near-native speaker, whose discourse may display less lexical density and less lexical sophistication than a native speaker (Waller 1993), and who may have problems with collocations or, for that matter, registers. As for oral teaching materials, they are often unauthentic (cf. Gilmore 2004 or Römer 2004), and the little authentic materials that are available to students in the classroom tend to take the form of broadcast news, that is, a very formal and carefully prepared type of speech, quite different from spontaneous conversations. The influence of speech on EFL writing therefore appears as a less convincing explanation and we may be slightly surprised that Lorenz (1999) offers it as his main argument to account for the spoken-like causal markers he observes in German learners’ writing. This does not mean, however, that the influence of speech should be ruled out

altogether. In fact, in some contexts, spoken English does have a place in learners' everyday life. Thus, in the Netherlands, students are exposed to spoken (or at least written-to-be-spoken)⁴ English through television programmes, which are subtitled, not dubbed (see de Haan 2002: 112). Even in countries where television programmes are dubbed, it looks as if new technologies like the Internet are making informal and more oral-like registers of English more accessible to young people. However, given that the learner populations represented in ICLE do not all live in an "English-friendly" environment and that the Web had little impact in the 1990s, when most of the ICLE subcorpora were collected, other explanations have to be sought for the spoken-like nature of learner academic writing.

4.2. L1 transfer

The influence of the mother tongue on learner language has been demonstrated by many studies, and there is no doubt, today, that "transfer is one of the major factors shaping the learner's interlanguage competence and performance" (Kohn 1986: 21). Yet, little has been said about the possible influence of L1 in terms of register. Paquot (in press), using the example of the first imperative plural form, shows that a phenomenon of "transfer of register" seems to be at work in the case of French-speaking learners. Figure 6 shows that *let's/let us* is more typical of speech than of writing. It also shows that, while it is overused by learners from most mother tongue backgrounds, this overuse is particularly striking among French-speaking learners, who use the imperative even more than in native speech.

Figure 6 about here

Figure 7, which gives the relative frequency of first imperative plural verbs in French school essays, French editorials and comparable data in English,⁵ suggests that this overuse of *let's/let us* in French EFL learners' essays may well be the result of L1 transfer, since imperatives appear to be much more frequent in written French, where they are used with many different verbs to structure discourse, e.g. (14) to (16). In fact, a comparison of Figures

⁴ "Written-to-be-spoken" refers for example to conversations in TV series, which are close to spontaneous speech, but still differ from it in a number of respects.

⁵ The French school essays come from the *Corpus de Dissertations Françaises* (CODIF), while the English school essays come from the LOCNESS corpus (see later). The French and English editorials are part of the TRILLED corpus (*TRILingual Louvain corpus of Editorials*).

6 and 7 reveals that the relative frequency of the first imperative plural form in native French is very similar to that of *let's/let us* in the French component of ICLE, which confirms the hypothesis of a transfer of register.

Figure 7 about here

- (14) *Prenons l'exemple des sorciers ou des magiciens au Moyen Age.* (CODIF)
'Let us take the example of wizards or magicians in the Middle Ages'
- (15) *Envisageons tout d'abord la question économique.* (CODIF)
'Let us first consider economic issues'
- (16) *Pensions, par exemple, à l'Espagne, qui, pendant quatre à huit siècles, a appris à côtoyer les peuples arabes.* (CODIF)
'Let us think, for example, about Spain which, for four to eight centuries, learnt to live with the Arab people'

However, the very fact that the imperative form is also overused by learners from other mother-tongue backgrounds which do not make use of first person plural imperative verbs to organise written academic discourse (e.g. Dutch, with its very low frequency of 16.3 imperatives pmw in data from the TRILLED corpus) proves that transfer cannot be the sole explanation and that other factors must also have a role to play (cf. Paquot 2007). Among these are teaching-induced and developmental factors, which are described in the next two sections.

4.3. Teaching-induced factors

Sometimes, the emergence of spoken-like features seems to be due to the teaching process itself. Most often recognised in the literature is the pernicious influence of undifferentiated lists of connectors found in textbooks, as in the following extract, taken from Jordan (1999):

Figure 8 about here

Such lists have been criticised for giving the misleading impression that the connectors are synonymous (see e.g. Crewe 1990: 317-318; Milton & Tsang 1993: 231-232). They have also been criticised for giving the impression that connective devices appear in sentence-initial

position only (Milton 1999). An additional problem with these lists is that the connectors are not differentiated in terms of register:

Students are drilled in the categorical use of a short list of expressions – often those functioning as connectives or alternatively those which are colourful and complicated (and therefore impressive) – regardless of whether they are used primarily in spoken or written language (if indeed at all), or to which text types they are appropriate (Milton 1998: 190).

With no explanation to resort to, the learner is likely to deduce that, say, *so* is just as appropriate as *consequently* or *therefore* in an academic essay, while it is in fact more characteristic of speech.

The teaching-induced factor is also at work in lists of translational equivalents such as those found in bilingual dictionaries. Below are some of the expressions found in the “essay writing” section of the French-English bilingual dictionary *Robert & Collins Senior* (1998):

- (17) Il était très mécontent, aussi a-t-il démissionné
= ... and so he ...
- (18) Prenons comme point de départ...
= Let us take ... as a starting point
- (19) N’oublions pas que...
= Let us not forget that...

The use of the connector *so* to express a consequence is more typical of spoken interactions than of academic prose, as noted earlier. As for the first imperative plural form, it is much less common in academic writing than in speech (see Figure 6) and its use in academic writing is lexically restricted to a few verbs such as *consider* or *examine* (see Swales et al. 1998). In the *Robert & Collins Senior* dictionary, by contrast, French imperatives are almost systematically translated into English by means of the *let us* form, no matter what the verb is. Such translations, presumably, distort learners’ sense of what is stylistically appropriate or not.

4.4. Developmental factors

A fourth explanation for the presence of spoken-like features in learner writing, which may go unnoticed in a comparison such as this one, using expert writing as a point of comparison, is

the influence of what may be referred to as developmental factors. In other words, some of the features highlighted above are not (solely) due to the non-nativeness of the subjects, but to the fact that they are novice writers, who have not quite acquired the rules of academic writing yet and lack knowledge of more formal alternatives to structure their discourse.

In order to study the role played by developmental factors, we compared the learner data with similar data produced by native students (i.e. novice writers). The corpus we used is the Louvain Corpus of Native English Essays (LOCNESS; see Granger 1996), which contains about 300,000 words of academic writing produced by British and American students. The analysis of LOCNESS reveals that, while native students do better than non-native students with respect to lexico-grammar and phraseology, they share learners' problem with register to a certain extent, overusing items which are more typical of speech than of writing. When compared with learner writing, (expert) academic writing and speech, novice writing often appears to occupy an intermediate position between academic writing and learner writing, with speech coming last. This cline is very clear in Figures 9 and 10, which present the relative frequencies of *maybe* and *I think*, respectively, in the four corpora. Both native and non-native students overuse these spoken-like items, although the overuse is more marked in learner writing than in novice writing. In addition, preliminary analyses on different components of LOCNESS point to the existence of a cline among native students, with A-level students favouring spoken-like features more than university students. All this suggests that an oral tone in writing is not limited to foreign learners, but is actually very much part of the process of becoming an expert writer.

Figure 9 about here

Figure 10 about here

5. Conclusion

The aim of this article was to examine the written production of upper-intermediate to advanced foreign learners of English and to situate it along the speech-writing continuum. We saw that learner academic writing, whatever the learner's mother tongue, exhibits many characteristics of speech, not only in terms of overuse or underuse of some items, but also with respect to the positioning of certain words. Several explanations may be put forward to

account for learners' "chatty" style. These include the influence of speech, L1 transfer, lack of knowledge of more formal alternatives, teaching-induced and developmental factors. It should be emphasised, however, that it is often difficult to pinpoint which factor is responsible for a particular spoken-like characteristic. Most of the time, it seems more reasonable to invoke a "complex interplay of factors" (Granger 2004: 135), with each factor reinforcing the effect of the other factors. If we want to gain more insight into the spoken-like nature of learner writing, it will therefore be necessary to carry out studies integrating these different factors, not as independent, but as interrelated variables.

Another possible improvement would be the use of corpora representing an even wider range of genres and text types than were available for this study. In particular, there is an urgent need for learner corpora which represent academic text types other than argumentative essays (cf. projects such as the British Academic Written Corpus and the Michigan Corpus of Upper-level Student Papers). Such corpora would make it possible to control for task type effects, better than was the case in this study. Learner corpus research would also greatly benefit from the design of comparable corpora of L1 and L2 writing produced by the same learners, since one cannot reasonably expect EFL learners to produce good academic texts in English if they are not able to perform such a complex task in their mother tongue in the first place.

Such improvements (and others) are necessary if we want to take studies like this one a step further and gain a better understanding of learners' attraction towards speech. We will then be ready for the final step, namely the preparation of appropriate remedial materials that will help learners overcome register-related problems.

Acknowledgements

Gaëtanelle Gilquin acknowledges the support of the Belgian National Fund for Scientific Research (F.N.R.S.). Magali Paquot acknowledges the support of the Communauté française de Belgique, which funded this research within the framework of the 'Action de recherche concertée' project entitled 'Foreign Language Learning: Phraseology and Discourse' (No. 03/08-301).

References

- Ädel, A. (2006) *The use of metadiscourse in argumentative texts by advanced learners and native speakers of English*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Ädel, A. (forthcoming) 'Involvement features in writing: Do time and interaction trump register awareness?'. In G. Gilquin, M.B. Diez-Bedmar and S. Papp (eds) *Linking up contrastive and learner corpus research*. Amsterdam & Atlanta: Rodopi.
- Aijmer, K. (2002) 'Modality in advanced Swedish learners' written interlanguage'. In S. Granger, J. Hung and S. Petch-Tyson (eds) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [Language Learning and Language Teaching 6] (pp. 55-76). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Altenberg, B. and M. Tapper (1998) 'The use of adverbial connectors in advanced Swedish learners' written English'. In S. Granger (ed.) *Learner English on Computer* (pp. 80-93). London & New York: Addison-Wesley Longman.
- Biber, D. (1988) *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Blanpain, K. (2006) *Academic Writing in the Humanities and Social Sciences: A Resource for Researchers*. Leuven: Acco.
- Carter, R. and M. McCarthy (1995) 'Grammar and the spoken language'. *Applied Linguistics* 16(2): 141-158.
- Conrad, S. (1999) 'The importance of corpus-based research for language teachers'. *System* 27: 1-18.
- Crewe, W.J. (1990) 'The illogic of logical connectives'. *ELT Journal* 44(4): 316-325.
- Cutting, J. (2000) 'Written errors of international students and English native speaker students'. In G.M. Blue, J. Milton and J. Saville (eds) *Assessing English for Academic Purposes* (pp. 97-113). Frankfurt am Main: Peter Lang.
- De Cock, S. (2003) *Recurrent sequences of words in native speaker and advanced learner spoken and written English: a corpus-driven approach*. Unpublished PhD thesis. Louvain-la-Neuve: Université catholique de Louvain.
- De Haan, P. (2002) 'The status of English in The Netherlands'. In S. Granger, E. Dagneaux and F. Meunier (eds) *The International Corpus of Learner English. Handbook and CD-ROM. Version 1.1* (pp. 111-115). Louvain-la-Neuve: Presses universitaires de Louvain. Available from <http://www.i6doc.com>.

- Field, Y. and L.M.O. Yip (1992) 'A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English'. *RELC Journal* 23(1): 15-28.
- Flowerdew, L. (2001) 'The exploitation of small learner corpora in EAP materials design'. In M. Ghadessy and R. Roseberry (eds) *Small Corpus Studies and ELT* (pp. 363-379). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Gilmore, A. (2004) 'A comparison of textbook and authentic interaction'. *ELT Journal* 58: 363-374.
- Gilquin, G., S. Granger and M. Paquot (2007) 'Writing sections'. In M. Rundell (Editor in chief) *Macmillan English Dictionary for Advanced Learners* (second edition) (pp. IW1-IW29). Oxford: Macmillan Education.
- Gilquin, G., S. Granger and M. Paquot (in press) 'Learner corpora: the missing link in EAP pedagogy'. In P. Thompson (ed.) *Corpus-based EAP Pedagogy*. Special issue of *Journal of English for Academic Purposes*.
- Granger, S. (1996) 'From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora'. In K. Aijmer, B. Altenberg and M. Johansson (eds) *Languages in Contrast. Text-based Cross-linguistic Studies* [Lund Studies in English 88] (pp. 37-51). Lund: Lund University Press.
- Granger, S. (1997) 'On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared'. In J. Aarts, I. de Mönnink and H. Wekker (eds) *Studies in English Language and Teaching* (pp. 185-198). Amsterdam & Atlanta: Rodopi.
- Granger, S. (1998) 'Prefabricated patterns in advanced EFL writing: collocations and formulae'. In A.P. Cowie (ed.) *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S. (2003) 'The International Corpus of Learner English: A new resource for foreign language learning and teaching and second language acquisition research'. *TESOL Quarterly* 37(3): 538-546.
- Granger, S. (2004) 'Computer learner corpus research: current status and future prospects'. In U. Connor and T. Upton (eds) *Applied Corpus Linguistics: A Multidimensional Perspective* (pp. 123-145). Amsterdam & Atlanta: Rodopi.
- Granger, S., E. Dagneaux, F. Meunier and M. Paquot (forthcoming) *International Corpus of Learner English. Handbook & CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.

- Granger, S. and P. Rayson (1998) 'Automatic lexical profiling of learner texts'. In S. Granger (ed.) *Learner English on Computer* (pp. 119-131). London & New York: Addison Wesley Longman.
- Granger, S. and J. Thewissen (2005) 'Towards a reconciliation of a "Can Do" and "Can't Do" approach to language assessment'. Paper presented at the *Second Annual Conference of EALTA (European Association of Language Testing and Assessment)*, Voss, Norway, 2-5 June 2005.
- Granger, S. and S. Tyson (1996) 'Connector usage in the English essay writing of native and non-native EFL speakers of English'. *World Englishes* 15: 17-27.
- Hinkel, E. (2003) 'Adverbial markers and tone in L1 and L2 students' writing'. *Journal of Pragmatics* 35(7): 1049-1068.
- Hyland, K. (1994) 'Hedging in academic writing and EAP textbooks'. *English for Specific Purposes* 13(3): 239-256.
- Hyland, K. and J. Milton (1997) 'Qualification and certainty in L1 and L2 students' writing'. *Journal of Second Language Writing* 6(2): 183-205.
- Jordan, R.R. (1999) *Academic Writing Course. Study Skills in English. Third Edition*. Harlow: Longman.
- Kohn, K. (1986) 'The analysis of transfer'. In E. Kellerman and M. Sharwood Smith (eds) *Crosslinguistic Influence in Second Language Acquisition* (pp. 21-34). New York: Pergamon Press.
- Leech, G. (1998) 'Preface'. In S. Granger (ed.) *Learner English on Computer* (pp. xiv-xx). London & New York: Addison Wesley Longman.
- Lorenz, G. (1999) 'Learning to cohere: causal links in native vs. non-native argumentative writing'. In W. Bublitz, U. Lenk and E. Ventola (eds) *Coherence in Spoken and Written Discourse. How to create it and how to describe it* (pp. 55-75). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Milton, J. (1998) 'Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment'. In S. Granger (ed.) *Learner English on Computer* (pp. 186-198). London & New York: Addison Wesley Longman.
- Milton, J. (1999) 'Lexical thickets and electronic gateways: making text accessible by novice writers'. In C.N. Candin and K. Hyland (eds) *Writing: Texts, Processes and Practices* (pp. 221-243). London & New York: Longman.
- Milton, J. and E.S.C. Tsang (1993) 'A corpus-based study of logical connectors in EFL students' writing: directions for future research'. In R. Pemberton and E.S.C. Tsang (eds)

- Studies in Lexis. Working Papers from a Seminar* (pp. 215-246). Hong Kong: Language Centre, The Hong Kong University of Science and Technology.
- Narita, M. and M. Sugiura (2006) 'The use of adverbial connectors in argumentative essays by Japanese EFL college students'. *English Corpus Studies* 13: 23-42.
- Neff, J., F. Ballesteros, E. Dafouz, F. Martínez and J.P. Rica (2004) 'The expression of writer stance in native and non-native argumentative texts'. In R. Facchinetti and F. Palmer (eds) *English Modality in Perspective* (pp. 141-161). Frankfurt am Main: Peter Lang.
- Neff, J., F. Ballesteros, E. Dafouz, F. Martínez and J-P. Rica (2007) 'A contrastive functional analysis of errors in Spanish EFL university writers' argumentative texts: corpus-based study'. In E. Fitzpatrick (ed.) *Corpus Linguistics beyond the Word: Corpus Research from Phrase to Discourse* [Language and Computers 23] (pp. 203-225). Amsterdam: Rodopi.
- Paquot, M. (2007) *EAP vocabulary in native English and EFL learner writing: A phraseology-oriented approach*. Unpublished PhD thesis. Louvain-la-Neuve: Université catholique de Louvain.
- Paquot, M. (in press) 'Exemplification in learner writing: A cross-linguistic perspective'. To appear in S. Granger and F. Meunier (eds) *Phraseology: An Interdisciplinary Perspective*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Rayson, P. (2003) *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Unpublished PhD thesis. Lancaster: Lancaster University.
- Recski, L.J. (2004) 'Expressing standpoints in EFL written discourse'. *Revista Virtual de Estudos da Linguagem* 3. Available online at www.revelhp.cjb.net (last accessed on May 20, 2007).
- Robert & Collins Senior. (1998). *Dictionnaire français-anglais/anglais-français. Cinquième édition*. New York: Harper Collins.
- Römer, U. (2004) 'Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching'. In G. Aston, S. Bernardini and D. Stewart (eds) *Corpora and Language Learners* (pp. 151-168). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Schleppegrell, M.J. (1996) 'Conjunction in spoken English and ESL writing'. *Applied Linguistics* 17(3): 271-285.
- Swales J., U. Ahmad, Y. Chang, D. Chavez, D. Dressen and R. Seymour (1998) 'Consider this: the role of imperatives in scholarly writing'. *Applied Linguistics* 19(1): 97-121.
- Waller, T. (1993) 'Characteristics of near-native proficiency in writing'. In H. Ringbom (ed.) *Near-native Proficiency in English* (pp. 183-294). Åbo: Åbo Akademi University.

FIGURES AND TABLES

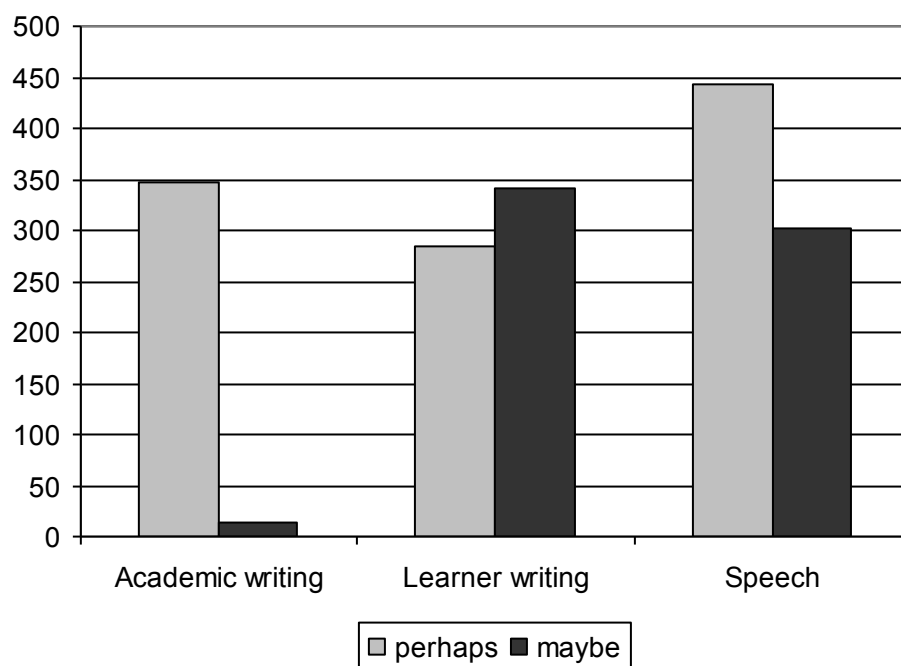


Figure 1. Relative frequency of *perhaps* and *maybe* in academic writing, learner writing and speech (relative frequency per million words)

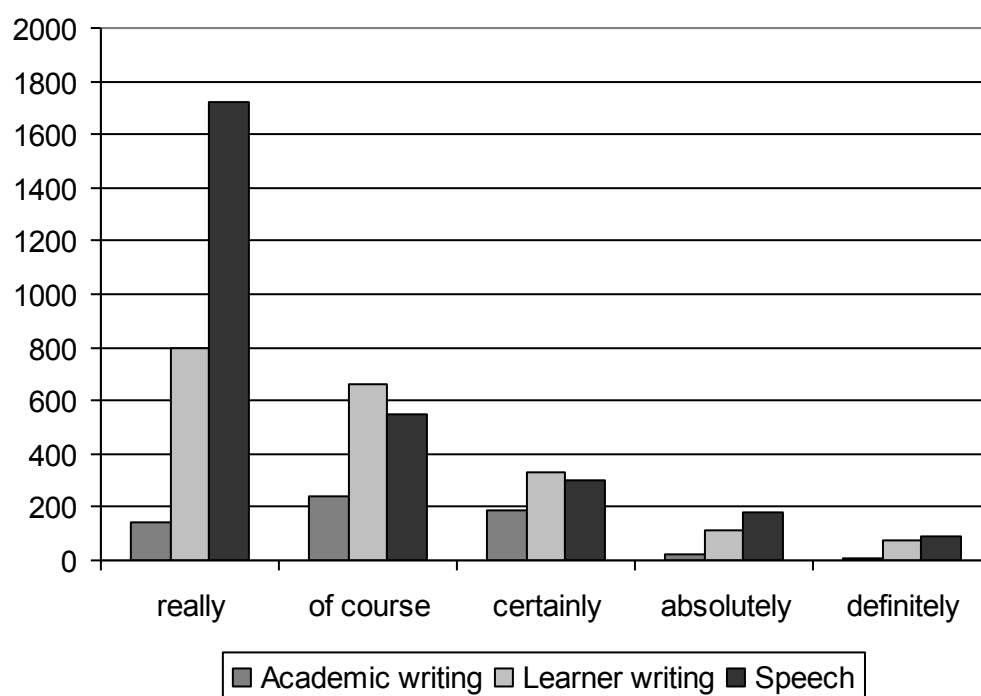


Figure 2. Relative frequency of amplifying adverbs in academic writing, learner writing and speech (relative frequency per million words)

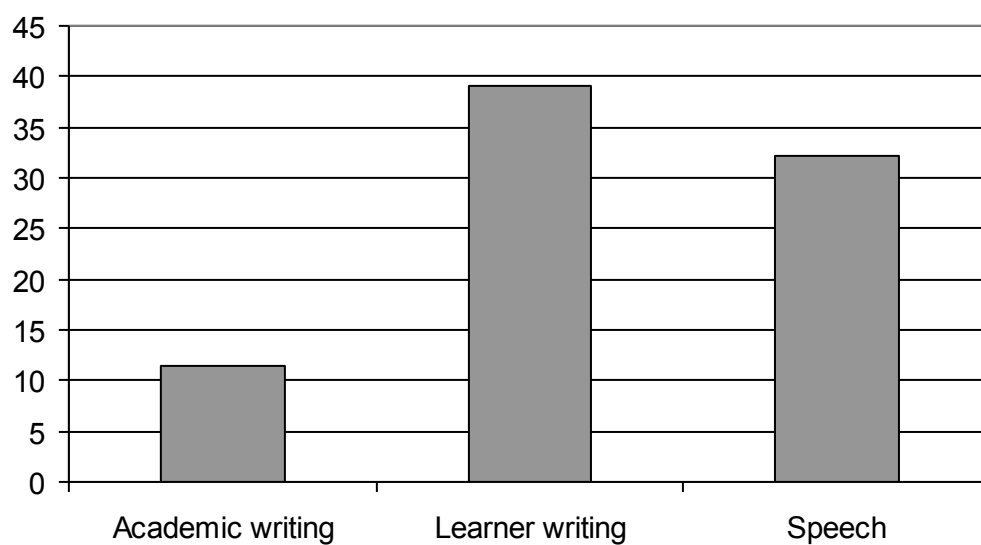


Figure 3. Relative frequency of *it seems to me* in academic writing, learner writing and speech (relative frequency per million words)

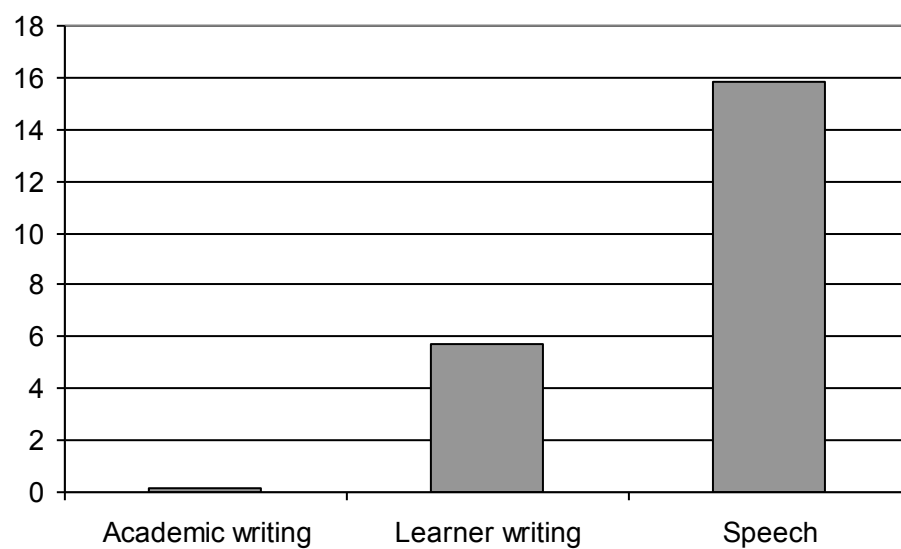


Figure 4. Relative frequency of *I would like / want / am going to talk about* in academic writing, learner writing and speech (relative frequency per million words)

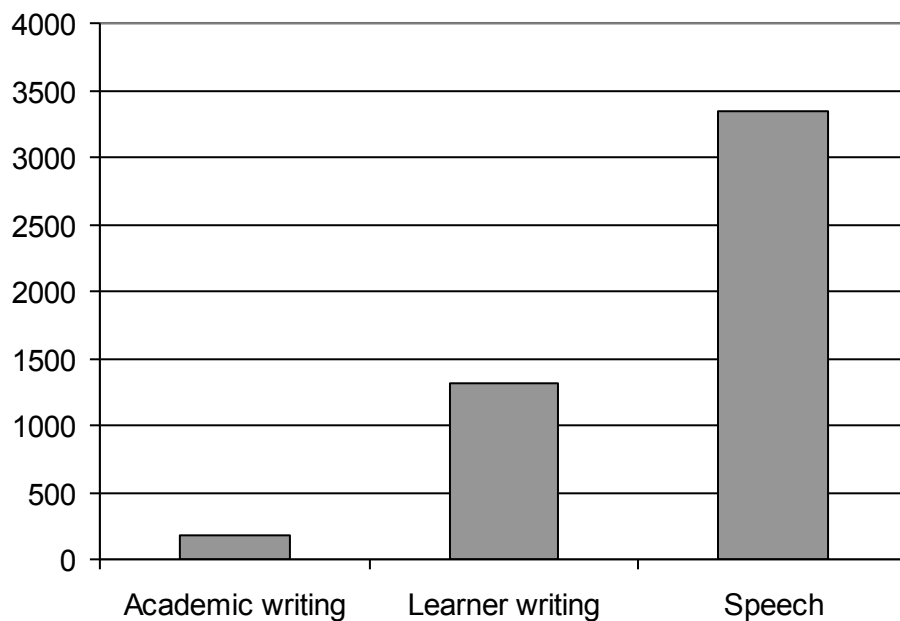


Figure 5. Relative frequency of sentence-initial *and* in academic writing, learner writing and speech (relative frequency per million words)

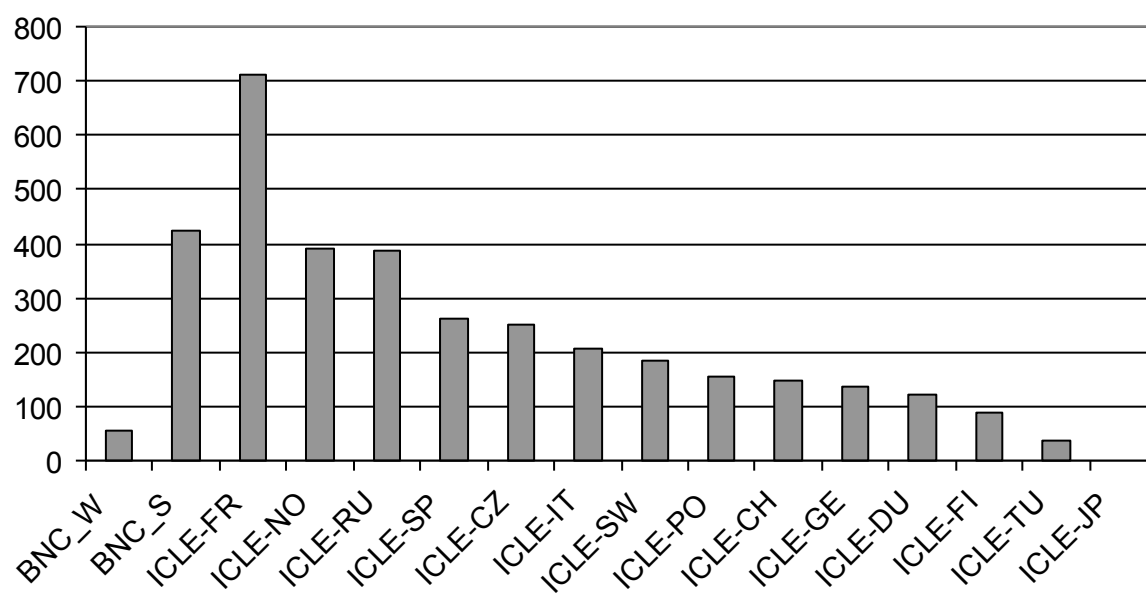


Figure 6. Relative frequency of *let's/let us* in academic writing, speech and learner writing from different L1 populations (relative frequency per million words)

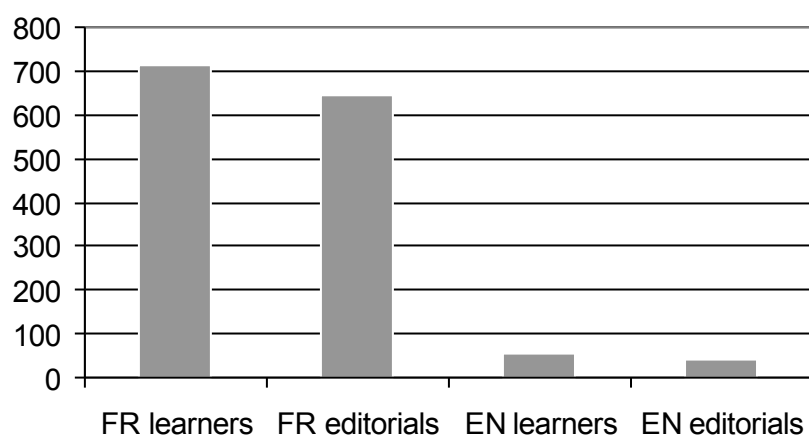


Figure 7. First imperative plural verbs in French and English (relative frequency per million words)

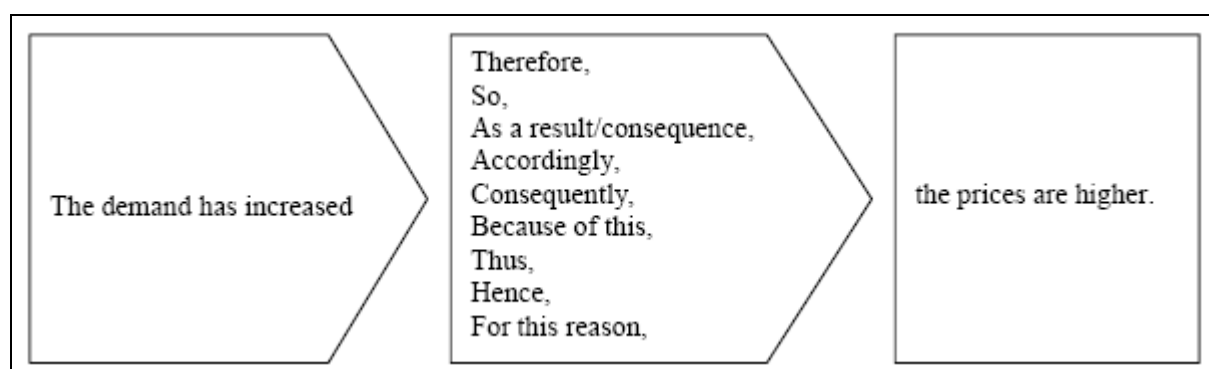


Figure 8. Connectives of result in Jordan (1999: 62)

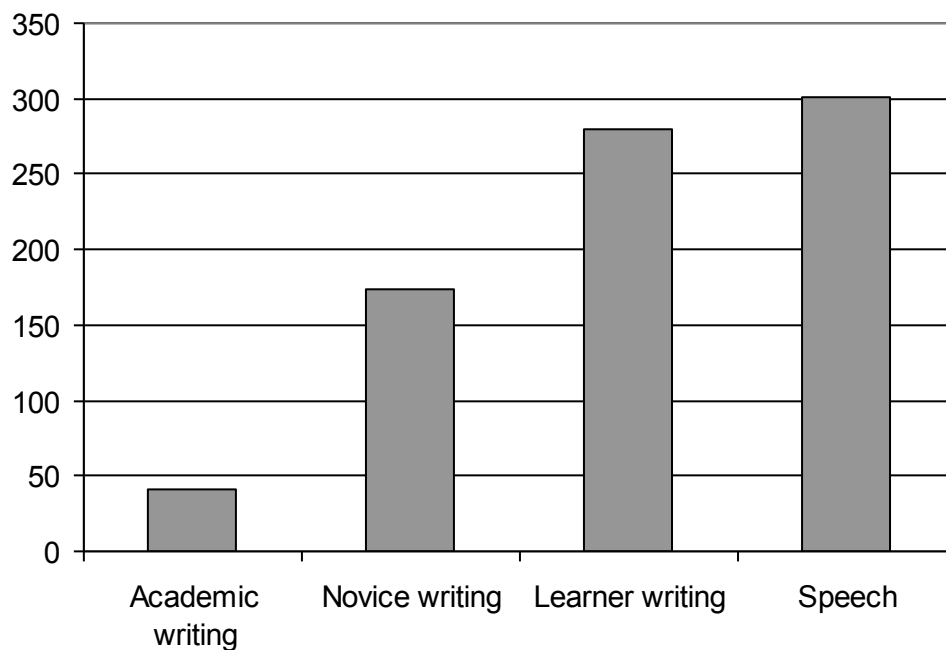


Figure 9. Relative frequency of *maybe* in academic writing, novice writing, learner writing and speech (relative frequency per million words)

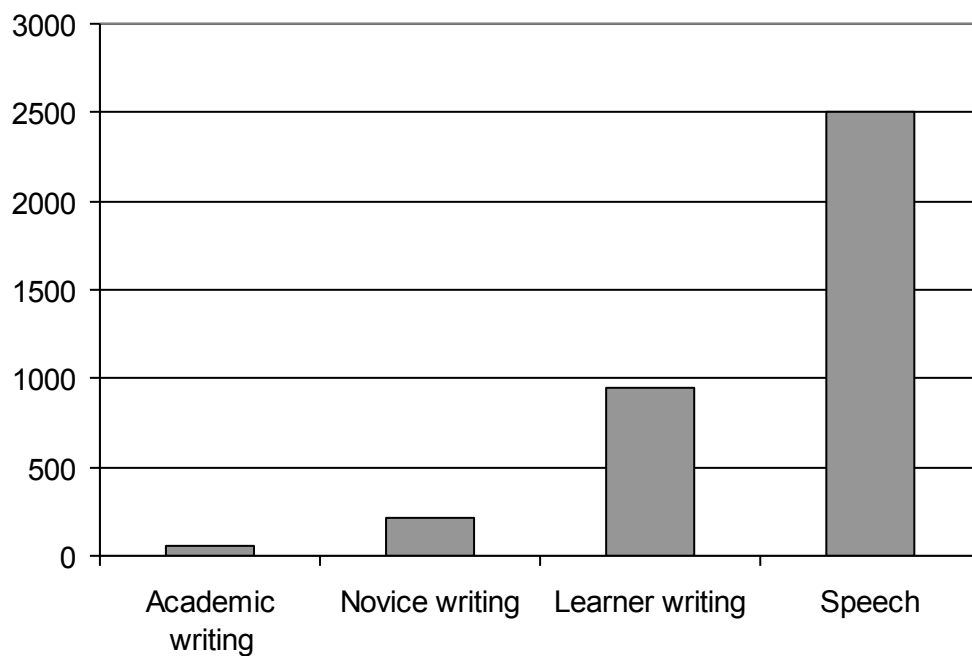


Figure 10. Relative frequency of *I think* in academic writing, novice writing, learner writing and speech (relative frequency per million words)

Nouns	alternative, approach, aspect, assumption, attempt, basis, comparison, conclusion, development, difference, emphasis, evidence, explanation, interest, issue, likelihood, point, problem, use, value, view
Verbs	achieve, adopt, allow, appear, argue, clarify, compare, concern, conduct, consist, contrast, deal, define, demonstrate, differ, distinguish, emerge, follow, indicate, interpret, present, provide, suggest, support
Adjectives	acceptable, adequate, central, common, considerable, crucial, detailed, different, distinct, effective, important, likely, major, possible, significant, specific, subsequent, systematic, unlike, valuable
Adverbs	also, closely, consequently, conversely, essentially, for example, for instance, however, in particular, largely, moreover, namely, partially, potentially, thereby, therefore, thus, widely
Others	according to, although, as opposed to, despite, given that, in addition to, in terms of, in the light of, provided, since, subject to, versus, whereas, whether

Table 1. Examples of EAP-specific words and phrases from Paquot's (2007) *Academic Keyword List*

Rhetorical function	Spoken-like overused lexical item
Exemplification	<i>like</i>
Cause and effect	<i>thanks to</i> <i>so</i> <i>because</i> <i>that/this is why</i>
Comparison and contrast	<i>look like</i> <i>like</i>
Concession	sentence-final adverb <i>though</i>
Adding information	sentence-initial <i>and</i> the adverb <i>besides</i>
Expressing personal opinion	<i>I think</i> <i>to my mind</i> <i>from my point of view</i> <i>it seems to me</i>
Expressing possibility and certainty	<i>really</i> <i>of course</i> <i>absolutely</i> <i>maybe</i>
Introducing topics and ideas	<i>I would like to/want/am going to talk about</i> <i>thing</i> <i>by the way</i>
Listing items	<i>first of all</i>

Table 2. Spoken-like overused lexical items per rhetorical function (Paquot 2007)