

# Spoken Features in Learner Academic Writing: Identification, Explanation and Solution

---

Gaëtanelle Gilquin<sup>1</sup> and Magali Paquot<sup>2</sup>

## Abstract

This paper shows, on the basis of a comparison of learner corpus data with written and spoken native corpus data, that learners tend to use spoken features when organising their academic writing. Possible explanations are offered to account for this problem of register confusion, including L1 transfer, L2 instruction and developmental factors. Suggestions are also made to help learners become more aware of register variation.

## 1. Introduction

The research reported in this paper originated in a large-scale project, whose aim was to study a number of rhetorical functions used to organise academic discourse, such as exemplification, reformulation, concession or the expression of a personal opinion. This project, which was undertaken in close collaboration with Macmillan Education for the second edition of the *Macmillan English Dictionary for Advanced Learners* (MED2), compared the use of such functions in native and non-native writing, and sought to identify the main difficulties experienced by learners of English when writing an academic essay (*cf.* Gilquin *et al.* forthcoming for a detailed description of the project and Gilquin *et al.* 2007 for the outcome of the project). One aspect which turned out to be problematic for learners from several mother tongue (L1) backgrounds was the use of a stylistically appropriate tone. More particularly, many learners use features which are more typical of speech than of writing, and which therefore give the essay an overly oral tone. In this paper, the spoken-like nature of learner academic writing is investigated through a three-fold process of identification (what are the most common spoken features present in learner writing?), explanation (how can the presence of such features be accounted for?) and solution (how can this lack of register-awareness be remedied?). Before turning to the results of the investigation, however, earlier studies in this area will be briefly considered, and more information will be given about the data and methodology used in this study.

## 2. Is L2 writing like English conversation?

This question, borrowed from the title of a talk given by William Crawford in 2005, has been answered positively by several linguists over the last few years. Thus, Crawford himself shows that features such as personal pronouns, contractions, the quantifier *all* or

---

<sup>1</sup> Fonds National de la Recherche Scientifique, Centre for English Corpus Linguistics, Université catholique de Louvain

*e-mail:* gaetanelle.gilquin@uclouvain.be

<sup>2</sup> Centre for English Corpus Linguistics, Université catholique de Louvain

*e-mail:* magali.paquot@uclouvain.be

the demonstrative pronoun *that*, which are markedly more frequent in conversation than in writing, tend to be used more by German, Spanish and Bulgarian learners than by native speakers. Similarly, Granger and Rayson (1998) demonstrate 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.*), but underuse many of the characteristics of formal writing, such as a high density of nouns and prepositions. Other studies have focused on more specific items, for example *I think* (Granger 1998, Aijmer 2002, Neff *et al.* 2007), *of course* (Granger and Tyson 1996, Altenberg and Tapper 1998, Narita and Sugiura 2006), *because* (Lorenz 1999) or *so* (Lorenz 1999, Anping 2002), showing that these items tend to be overused by learners and that this overuse gives learner writing a distinctly oral tone.

All these studies point to the same lack of register-awareness among (even advanced) learners of English and underline the overly oral tone of learner writing. They are, however, limited to a small range of learner L1 populations (usually just one, sometimes up to three, as in Crawford 2005). The present study, by contrast, examines the written production of learners from a large number of mother tongue backgrounds, in an attempt to uncover more general trends with respect to register variation.

### 3. Data and methodology

The learner data we analysed come from the second edition of ICLE, the *International Corpus of Learner English* (Granger *et al.*, forthcoming). This new edition contains over 3.5 million words of academic writing (about 6,000 essays) produced by relatively advanced foreign learners from sixteen mother tongue backgrounds belonging to different language families (e.g. Italian, German, Japanese, Finnish). Given the importance of settings for learner writing (*cf.* Ädel 2006), only those texts which were produced under the same conditions (untimed, no reference tools, argumentative essays) were included, which reduced the sample to fourteen L1 populations and some 1.5 million words (*cf.* Table 1). In the analysis of the results, no distinction will be made between the different components of the ICLE subcorpus, but features will only be mentioned if they are found in a majority of the L1 populations.

ICLE-COMPONENT	NUMBER OF WORDS
Chinese	27,451
Czech	130,768
Dutch	162,243
Finnish	125,292
French	136,343
German	109,556
Italian	47,739
Japanese	21,451
Norwegian	128,544
Polish	140,521
Russian	165,937
Spanish	99,119
Swedish	48,060
Turkish	105,006
<b>TOTAL</b>	<b>1,404,776</b>

**Table 1:** Breakdown of the word counts of the ICLE subcorpus

The data from ICLE were compared with two corpora of native English, representing written and spoken registers, namely the academic component of the *British National Corpus* (BNC), totalling fifteen million words, and its spoken component, with ten million words.

From these three subcorpora, a number of words and phrases were extracted which fulfil one of twelve rhetorical functions particularly prominent in academic writing.<sup>3</sup> These words and phrases, approximately 350 in total, formed the basis of our analysis. In the next section, we will present some of the most striking examples of learner behaviour coming closer to spoken English than to academic writing, we will give possible explanations for such behaviour and we will suggest some solutions to this problem.

#### 4. Rhetorical functions in learner writing: spoken features

##### 4.1. Identification

Most of the twelve rhetorical functions we examined turned out to be characterised in learners' essays by a number of items which are more typical of speech than of academic writing, as shown in Table 2. Items such as *thanks to*, *look like*, *maybe* and *by the way* are all overused by a majority of the learners in ICLE, as compared to native writers, and are all more common in the spoken component of the BNC than in the academic component.

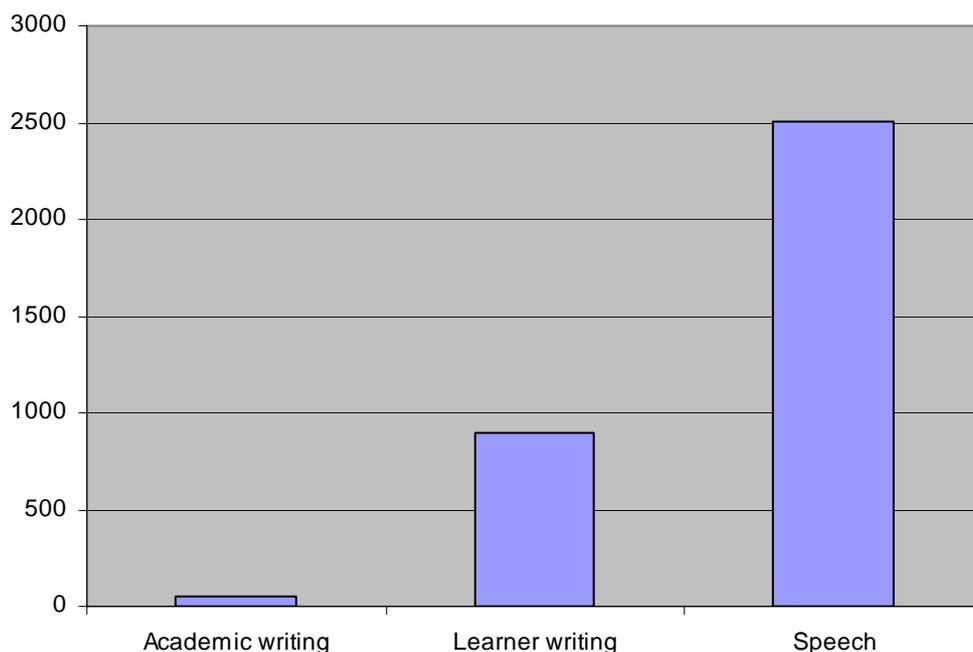
RHETORICAL FUNCTION	SPOKEN-LIKE OVERUSED LEXICAL ITEM
Exemplification	<i>like</i>
Cause and effect	<i>thanks to</i> <i>so</i> <i>because</i> <i>that/this is why</i>
Comparison and contrast	<i>look like</i> <i>like</i>
Concession	sentence-final adverb <i>though</i>
Adding information	sentence-initial <i>and</i> 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, forthcoming)

<sup>3</sup> These twelve functions are: (1) adding information; (2) comparing and contrasting: describing similarities and differences; (3) exemplification: introducing examples; (4) expressing cause and effect; (5) expressing personal opinions; (6) expressing possibility and certainty; (7) introducing a concession; (8) introducing topics and related ideas; (9) listing items; (10) reformulation: paraphrasing or clarifying; (11) reporting and quoting; (12) summarizing and drawing conclusions.

This list of spoken features reveals some interesting tendencies. First, it appears that learners overuse a number of expressions which make them particularly visible as writers, *cf. I think, to my mind, from my point of view* and *it seems to me* to express a personal opinion, and *I would like to/want/am going to talk about* to introduce a topic or an idea. Like all the features listed in Table 2, this is true irrespective of the learner’s mother tongue, even though there are some differences in frequency between the various L1 populations (*cf. Paquot, forthcoming*), which may be related to the different degrees of writer visibility typical of academic writing cross-linguistically, highlighted by e.g. Connor (1996) or Vassileva (1998) (*cf. Section 4.2 on the possible influence of L1*). Some examples of learners’ high visibility as writers are given in (1) to (3) and Figure 1 illustrates the situation for *I think*. While *I think* is rare in native academic writing (54.38 occurrences per million words), it is much more frequent in speech, with a relative frequency of 2,502 occurrences per million words. In comparison with native writing, learner writing exhibits a statistically significant overuse of the expression (893.38 occurrences per million words), which contributes to the oral tone of learners’ essays.

- (1) This is why *I think* that English is entitled to be called a world language. (ICLE-FI)
- (2) *From my point of view*, this is what the law should try to prevent. (ICLE-TU)
- (3) In this essay *I would like to talk about* the negative aspects of money. (ICLE-DU)

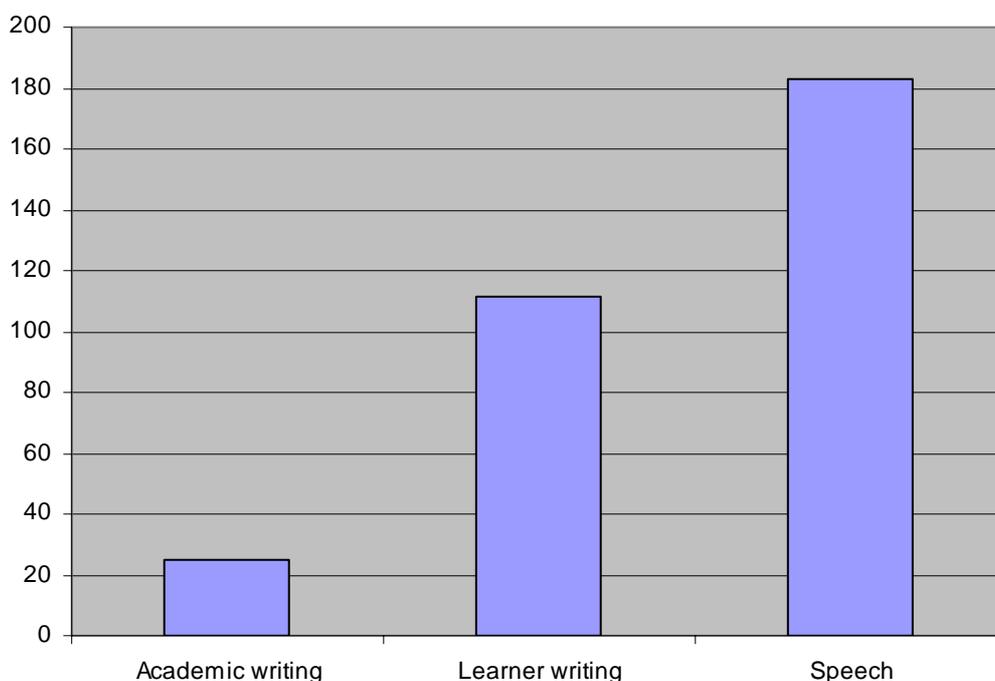


**Figure 1:** Relative frequency of *I think* in academic writing, learner writing and speech (relative frequency per million words)

Another characteristic of the list in Table 2 is the presence of emphasisers, especially in the form of adverbs expressing certainty, *cf. really, of course, absolutely* (4). The overuse of *first of all* (5) may also be seen in this light, since this expression is normally used to emphasise the first item of a list. It seems as if learners are so keen to get their message across that they put extra emphasis on it, thus creating an impression of “overstatement” (Lorenz 1998). This, however, marks their writing as spoken-like,

since emphasisers are more common in speech, as illustrated in Figure 2 for *absolutely*. Academic writing, by contrast, is much more reliant on hedges, which qualify, rather than emphasise, a statement.

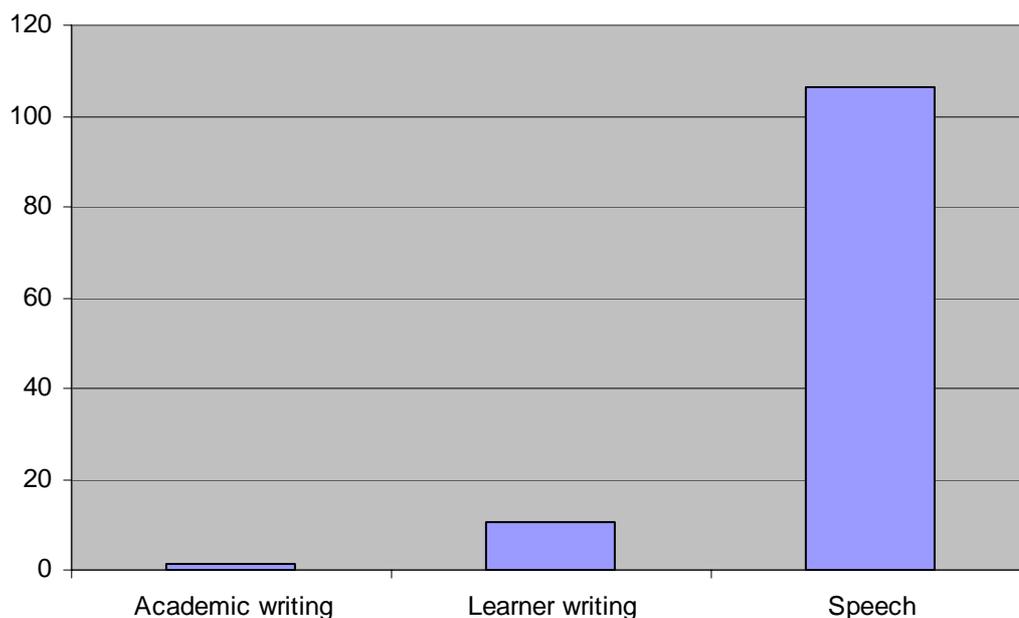
- (4) These verses make it *absolutely* clear that in the eyes of God there is nothing wrong with being rich. (ICLE-RU)
- (5) *First of all*, it is important to highlight the fact that the situation of women greatly improved this last century. (ICLE-FR)



**Figure 2:** Relative frequency of *absolutely* in academic writing, learner writing and speech (relative frequency per million words)

The list in Table 2 also contains two items whose position in the sentence is problematic, namely *though* (to express concession) and *and* (to add information). While the former is quite commonly found in academic writing as a conjunction, introducing a clause, it is extremely rare as a sentence-final adverb, a function which is significantly more frequent in learner writing and more typical of spoken English (*cf.* Figure 3 and example (6)). As for *and*, it is often used by learners in sentence-initial position, *cf.* (7), but this position is more characteristic of speech than of academic writing.

- (6) This doesn't mean that we no longer have a place for dreams, *though*. (ICLE-SW)
- (7) *And* finally there are a lot of films and programmes which keep you amused after your hard workday. (ICLE-PO)



**Figure 3:** Relative frequency of sentence-final *though* in academic writing, learner writing and speech (relative frequency per million words)

Finally, let us briefly mention that overuse of certain spoken-like items may go hand in hand with misuse. The adverb *besides* is a case in point. Not only is this spoken-like item overused by learners, but it is also misused. The typical function of *besides* in native English is to introduce a final point or argument that is decisive. In learner English, however, *besides* tends to be used simply to introduce an additional point, as in (8).

- (8) It is so sad that family members simply do not communicate. *Besides*, the happiness of the youngsters can also be affected by television. (ICLE-CH)

Now that some of the spoken features of learner writing have been identified, we can suggest possible explanations to account for learners' predilection for such features.

## 4.2. Explanation

One explanation that may immediately come to mind when trying to understand the spoken-like nature of learner writing is the influence of speech. This explanation has been suggested for learners of English as a Second Language, who mainly learn English through oral interaction (*cf.* Schleppegrell 1996 or Hinkel 2003). In the case of learners of English as a Foreign Language (EFL), however, this explanation seems less probable, given the limited spoken input to which learners are exposed. L2 classrooms may be seen as "impoverished learning environments" (Kasper 1997). Non-native teachers are at best near-native speakers, whose discourse often displays less lexical density and less lexical sophistication than a native speaker (*cf.* Waller 1993), and even native teachers tend to simplify language when addressing foreign learners (Henzl 1973). As for textbooks, they present learners with a variety of speech that is often quite different

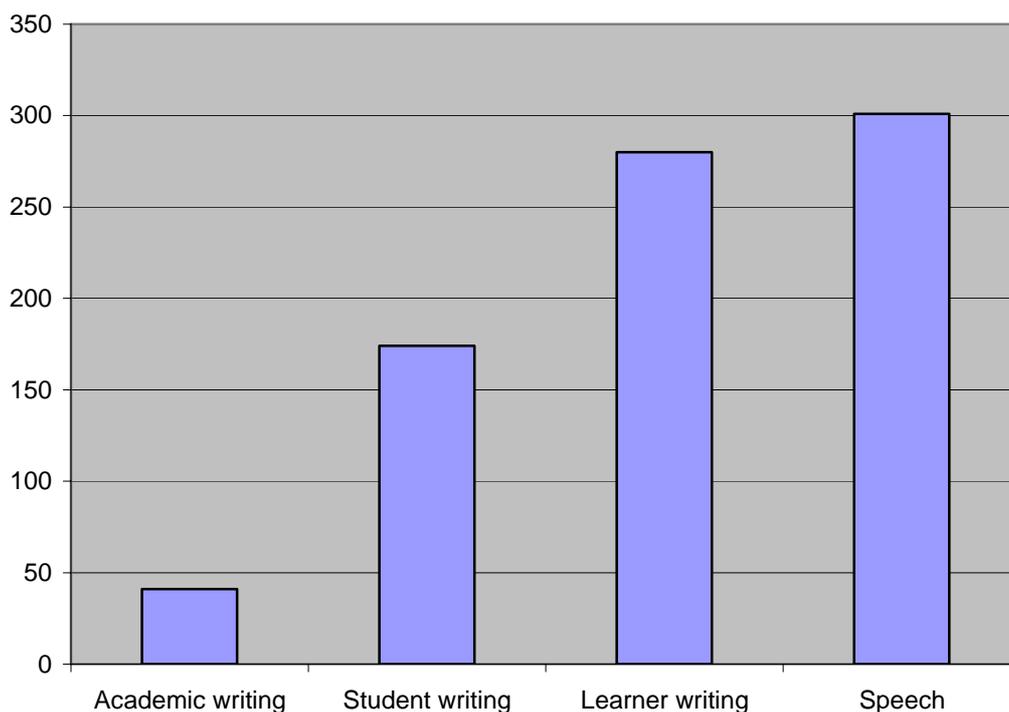
from authentic speech (*cf.* Gilmore 2004 or Römer 2004). Some learners are regularly exposed to authentic spoken English through the media (films, radio, internet, *etc.*) but it is not the majority of EFL learners, and certainly not of the EFL learners who contributed to ICLE in the 1990s. The influence of spoken registers, therefore, is unlikely to play an important role in the situation described in Section 4.1, and other explanations have to be found.

One such explanation is L2 instruction. Grammatical accuracy tends to be considered more important than stylistic appropriacy, and the materials to which learners have access is often inadequate. Especially harmful are the lists of connectors presented in textbooks with little context and no comments, as if these connectors were interchangeable (*cf.* Crewe 1990: 317-318, Milton and Tsang 1993: 231-232). These lists ignore any possible differences in register, with the result, for example, that learners will study *therefore* and *so* on a par with each other (Laws 1999: 55), not aware of the fact that the former is more appropriate than the latter to express consequence in an academic essay. In addition, L1-specific teaching materials may also reinforce transfer phenomena, as when the French-English bilingual dictionary *Robert and Collins Senior* (1998) recommends using the imperative to translate the French phrase *N'oublions pas que...* (*Let us not forget that...*), while the imperative is relatively infrequent in English academic writing (*cf.* below).

The examination of academic essays produced by native students brings to light another explanation for the spoken-like nature of learner writing, namely the influence of developmental factors. We compared the ICLE data with data from the Louvain Corpus of Native English Essays (LOCNESS, *cf.* Granger 1996), which contains about 300,000 words of academic writing produced by British and American students, and came to the conclusion that novice writers tend to use spoken features, regardless of whether English is their mother tongue or not. Thus, Figure 4, which gives the frequency of *maybe* in four varieties of English (academic writing, student writing, learner writing and speech), shows that native students also have a tendency to overuse this spoken-like adverb, although it is slightly less marked than among EFL learners. Register confusion, therefore, seems to be as much part of the process of acquiring a foreign language as it is part of the process of becoming an expert writer.

Finally, although in what precedes we considered features displayed by learners from many different mother tongue backgrounds, some influence of L1 is still at times noticeable. Paquot (in press) shows that, while the first imperative plural form *let's/let us*, which is more typical of speech than of academic writing, is overused by learners from most mother tongue backgrounds, this overuse is particularly striking among French-speaking learners. She explains this by a transfer of register and cites as evidence the almost identical frequency of the imperative in native French essays and in the English essays produced by French-speaking learners. A phenomenon of transfer is also quite clear in the frequent use, among French-speaking learners, of the expression *let us take the example of*, which has a literal equivalent in French, *prenons l'exemple de*.

Now that possible explanations for learners' behaviour have been highlighted, it becomes easier to see how the situation may be remedied. This is what we discuss below.



**Figure 4:** Relative frequency of *maybe* in academic writing, (native) student writing, learner writing and speech (relative frequency per million words)

### 4.3. Solution

Given learners' lack of register-awareness, it is important to include in the curriculum consciousness-raising activities, aimed at underlining the differences that exist between written and spoken registers. This could involve, for example, pointing out to the students that in (9), the phrase *to me* gives an oral tone to the sentence and that it can easily be left out without obscuring the fact that the author is expressing a personal opinion. Rewriting exercises could also be proposed to the students, with the instructions to rewrite a short paragraph in a more "academic-like" fashion. In (10), this would mean, among others, replacing the sentence-final adverb *though* by another expression such as *however*, expanding the contractions and getting rid of the phrases *I think* and *according to me*. By using corpus examples, as is the case of (9) and (10), one would ensure that the problems presented to the students are authentic problems. Note that, because spoken features are also, though to a lesser extent, characteristic of novice native writing, such activities could benefit native students too.

- (9) It seems to me that what these characters share is a solitude born out of love.
- (10) This isn't the right way, though. I think being able to talk about the problems in your couple and trying to find solutions to go on living together is the better way. Divorce isn't the solution, according to me.

In L1 homogeneous classrooms, it is possible for the teacher to adopt a more L1-specific approach. Going back to the example of the imperative, French-speaking learners could be shown that the form *let's/let us* is not very common in English academic writing, unlike the imperative in French. They could also be presented with

alternatives to express the meaning of the imperative (e.g. *This may be illustrated by...* instead of *Let us take the example of...*). With more research on the exact role of L1, it would be possible to pinpoint more precisely the spoken features that are worth considering in a particular learner population.

Finally, consciousness-raising and other register-related activities should ideally extend beyond the classroom and find their way into pedagogical materials. This is one of the goals of the project undertaken in collaboration with Macmillan Education and referred to earlier. This project has resulted in corpus-based writing sections (Gilquin *et al.* 2007) in MED2. While the sections contain general advice about how to express several important rhetorical functions in academic writing and how to avoid common errors, it also includes information on register, and more particularly, on spoken features which tend to be overused by learners. Figure 5 illustrates the treatment of the noun *thing* in the writing sections. In the form of a “Be careful!” note, it draws the reader’s attention to the stylistic properties of the vague noun *thing* and underlines its spoken-like character. The note also contains an authentic learner sentence and proposes a possible improvement of the sentence by means of the use of the noun *question*. Learners can also visualise the register information through the graph comparing academic writing and speech. Because MED2 is intended for readers from several mother tongue backgrounds, the writing sections do not include L1-specific register problems. However, more research in this area could lead to L1-specific materials suitable for the teaching of stylistic aspects such as the ones discussed here.

**BE CAREFUL!** Many learners use the noun *thing* to introduce a new topic, but this noun is more typical of speech and informal writing. In academic writing and professional reports, it is important to be as precise as possible, and you should avoid vague nouns such as *thing*.

? *The first thing to be discussed is: who has the right to adopt children?*

✓ *The first question to be addressed is: who has the right to adopt children?*

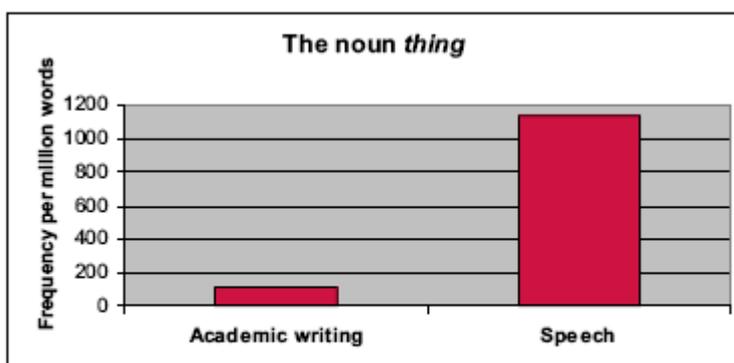


Figure 5: Treatment of the noun *thing* in MED2 (Gilquin *et al.* 2007: IW22)

## 5. Conclusion

This paper has brought to light a general tendency, among advanced learners of English from a wide range of mother tongue backgrounds, to use in their written production

words and phrases which are more typical of speech than of academic writing. Several possible explanations have been offered to account for this phenomenon, including L1 influence, L2 instruction and developmental factors. Consciousness-raising activities have also been proposed that should help learners become more aware of register variation and of the importance of adopting a stylistically appropriate tone in academic writing. In addition, a sample from a writing section included in the second edition of the *Macmillan English Dictionary for Advanced Learners* has been presented which addresses, among others, problems of register. Because such concrete achievements are still few and far between, however, many learners around the world still have to rely on less than perfect teaching materials. It is to be hoped that the avenues we have explored here will inspire other researchers to delve more deeply into this issue and propose their own remedial work.

## References

- Ädel, A. (2006) *The Use of Metadiscourse in Argumentative Texts by Advanced Learners and Native Speakers of English*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- 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*, pp. 55–76. Amsterdam and 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 and New York: Addison-Wesley Longman.
- Anping, H. (2002) On the discourse marker *so*, in P. Peters, P. Collins and A. Smith (eds) *New Frontiers of Corpus Research*, pp. 41–52. Amsterdam: Rodopi.
- Connor, U. (1996). *Contrastive Rhetoric: Cross-cultural Aspects of Second Language Writing*. New York: Cambridge University Press.
- Crawford W. (2005) Is L2 writing like native-English conversation? Paper presented at ICAME 26 (International Computer Archive of Modern and Medieval English) – AAACL 6 (American Association of Applied Corpus Linguistics), University of Michigan, 12–15 May 2005.
- Crewe, W.J. (1990) 'The illogic of logical connectives'. *ELT Journal* 44, 316–25.
- Gilmore, A. (2004) 'A comparison of textbook and authentic interaction'. *ELT Journal* 58, 363–74.
- 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 (forthcoming) 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*, pp. 37–51. Lund: Lund University Press.
- Granger, S. (1998) Prefabricated patterns in advanced EFL writing: collocations and formulae, in A.P. Cowie (ed.) *Phraseology: Theory, Analysis and Applications*, pp. 145–60. Oxford: Oxford University Press.

- Granger, S., E. Dagneaux, F. Meunier and M. Paquot (forthcoming) International Corpus of Learner English. Handbook and CD-ROM. Version 2. Louvain-la-Neuve: Presses universitaires de Louvain.
- Granger, S. and S. Petch-Tyson (1996) 'Connector usage in the English essay writing of native and non-native EFL speakers of English'. *World Englishes* 15, 19–29.
- Granger, S. and P. Rayson (1998) Automatic lexical profiling of learner texts, in S. Granger (ed.) *Learner English on Computer*, pp. 119–31. London and New York: Addison Wesley Longman.
- Henzl, V. (1973) 'Linguistic register of foreign language instruction'. *Language Learning* 23, 207–222.
- Hinkel, E. (2003) 'Adverbial markers and tone in L1 and L2 students' writing'. *Journal of Pragmatics* 35, 1049–68.
- Kasper, G. (1997) *Can pragmatic competence be taught?* Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center. Available on-line from <http://www.nflrc.hawaii.edu/NetWorks/NW06> (accessed: 18 June 2007).
- Laws, A. (1999) *Writing Skills*. Oxford: Summertown Publishing.
- Lorenz, G. (1998) Overstatement in advanced learners' writing: stylistic aspects of adjective intensification, in S. Granger (ed.) *Learner English on Computer*, pp. 53–66. London and 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 and Philadelphia: John Benjamins Publishing Company.
- 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. 215–46. 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 (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*, pp. 203–225. Amsterdam: Rodopi.
- 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 and Philadelphia: John Benjamins Publishing Company.
- Paquot, M. (forthcoming) *EAP vocabulary in native English and EFL learner writing: A phraseology-oriented approach*. PhD Thesis. Louvain-la-Neuve: Université catholique de Louvain.
- Robert and 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. 151–68. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Schleppegrell, M.J. (1996) 'Conjunction in spoken English and ESL writing'. *Applied Linguistics* 17, 271–285.

- Vassileva, I. (1998) 'Who am I/who are we in academic writing? A contrastive analysis of authorial presence in English, German, French, Russian, and Bulgarian'. *International Journal of Applied Linguistics* 8, 163–90.
- 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.