

Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora

Magali Paquot¹, Hilde Hasselgård² and Signe Oksefjell Ebeling²

¹Centre for English Corpus Linguistics, Université catholique de Louvain

²University of Oslo

Abstract

Previous studies have shown that learner writing is often characterized by a more involved style than the writing of their native peers, as evidenced by a high number of writer/reader (W/R) visibility features such as first and second person pronouns, *let's* imperatives, epistemic modal adverbs (e.g. *certainly*, *maybe*) and questions (cf. e.g. Petch-Tyson 1998, Altenberg and Tapper 1998). The aim of this study is to analyse French and Norwegian learners' use of W/R visibility features across genres, comparing argumentative texts from the *International Corpus of Learner English* (ICLE) with discipline-specific texts from the *Varieties of English for Specific Purposes dAtabase* (VESPA), to investigate whether learners are generally more overtly present within their academic writing or whether the features commonly attributed to EFL learners' involved style are prompted by the argumentative type of texts that has usually been analysed in learner corpus research. Results show that, when compared to native speakers' writing within the same discipline, texts produced by French and Norwegian learners display an overuse of W/R visibility features. There are, however, generally fewer features of W/R visibility in the discipline-specific texts, thus suggesting that learners adapt to genre requirements to some extent.

Keywords: writer/reader visibility, genre, argumentative essays, ESP, discipline-specific writing

1. Introduction

Features of writer/reader (W/R) visibility are used “to express personal feelings and attitudes and to interact with readers” (Petch-Tyson, 1998: 108) and include first and second person reference, speaker's mental processes, monitoring of information flow, emphatic particles, evaluative modifiers, imperatives and questions. Learner corpus-based studies have shown that learner writing, whatever the learners' mother tongues, exhibits many features of W/R visibility. For example, Granger and Rayson (1998) have shown that French-speaking learners overuse first and second person pronouns. Virtanen (1998) has reported on EFL learners' overuse of direct questions. EFL learners' heavy reliance on *I think* to express their personal opinion is reported by Granger (1998) for French L1, Neff et al. (2007) for Spanish L1, Aijmer (2001, 2002) for Swedish L1, and Hasselgård (2009) for Norwegian L1; Japanese, French and Swedish learners' overuse of *of course* is highlighted by Narita and Sugiura (2006), Granger and Tyson (1996) and Altenberg and Tapper (1998) respectively.

Several explanations have been put forward to account for EFL learners' involved style. These include the influence of speech, L1 transfer, aspects of teaching, and cultural and developmental factors (Petch-Tyson 1998; Gilquin and Paquot, 2008; Paquot, 2010). As noted by Recski (2004), however, in the case of argumentative essays such as those contained in the *International Corpus of Learner English* (ICLE), "personal references and subjective attitudes are certainly hard to avoid", as learners are explicitly prompted to give their personal opinions. ICLE prompts 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 his novel *Animal Farm*, George Orwell wrote 'All men are equal but some are more equal than others'. How true is this today?"

The aim of this study is to compare features of W/R visibility in learner writing across genres to investigate whether learner writers are generally more overtly present within their academic writing or whether the features commonly attributed to EFL learners' involved style are prompted by the argumentative type of texts that has usually been analysed in learner corpus research. Ädel (2008) found that the presence of involvement features in learner writing correlates with task setting (including conditions of writing such as the timed/untimed factor). We hypothesise that genre will be another influential factor and that the more academic nature of linguistics papers will be less welcoming to such features. Petch-Tyson (1998) showed that Scandinavian (i.e. Swedish and Finnish) learners of English have a higher degree of W/R visibility in their argumentative texts than French learners. We will thus also explore the extent to which the two learner groups differ in a similar way when writing in an academic discipline.

2. Material and methodology

We compare features of W/R visibility in the French and Norwegian sub-components of the *International Corpus of Learner English* (ICLE)¹ (Granger et al., 2009) and the *Varieties of English for Specific Purposes dAtabase* (VESPA).² ICLE consists of essays written within the framework of an English language course by intermediate to advanced EFL learners of different mother tongue backgrounds. The French (ICLE-FR) and Norwegian (ICLE-NO) sub-corpora used in this study were compiled on the basis of one learner variable (i.e. the mother tongue variable, either French or Norwegian) and three text variables, i.e. all texts are *untimed argumentative* essays potentially written with the help of *reference tools*.

The VESPA learner corpus project, on the other hand, aims at building a large corpus of English for Specific Purposes (ESP) texts written by L2 writers from various mother tongue backgrounds. The corpus currently contains papers, reports, etc. produced by BA and MA students in the context of a variety of content courses in linguistics, business, medicine, and engineering. The French and Norwegian components of VESPA investigated here are made up of essays and term papers written for linguistics courses (e.g. morphology, syntax, lexicology, corpus linguistics) taught in English. The criteria used for the ICLE subsets were also applied for the VESPA sub-corpora, i.e. the texts were written by students with either French or

¹ <http://www.uclouvain.be/en-cecl-icle.html>

² <http://www.uclouvain.be/en-cecl-vespa.html>

Norwegian as their L1, the writing of the assignments was not timed, and the students did have access to reference tools.

The study also draws on corpora of native student writing for comparison. The *Louvain Corpus of Native English Essays* (LOCNESS)³ corpus is used as a control corpus for the argumentative writing represented in the ICLE sub-corpora, while the *British Academic Written English* (BAWE)⁴ corpus is used as comparable data for the discipline-specific (linguistics) VESPA corpus.

The VESPA and BAWE corpora are annotated with TEI-conformant tags and features such as direct quotes and examples are tagged (when they have been explicitly marked as such by the students). This makes it possible to exclude text that is clearly not the students' own from the analysis.⁵ The VESPA-NO sub-corpus, for example, contains 298,000 words. When annotated material is excluded, the total number of words is reduced to 221,669 (cf. Table 1).

	Number of texts	Number of words
ICLE-FR	230	137,472
ICLE-NO	189	128,093
VESPA-FR	119	359,990
VESPA-NO	198	221,669
LOCNESS	88	100,702
BAWE-LING ⁶	76	195,606

Table 1. Overview of the corpora in terms of number of texts and words

The selection of corpora listed in Table 1 allows for a range of comparisons, addressing the various research questions outlined in Section 1. Our investigation involves comparisons along several dimensions:

- argumentative vs. more academic-like (i.e. discipline-specific) learner writing (ICLE vs. VESPA)
- non-native (i.e. learners) vs. native speakers of English (ICLE vs. LOCNESS; VESPA vs. BAWE)
- French-speaking vs. Norwegian learners (ICLE-FR vs. ICLE-NO; VESPA-FR vs. VESPA-NO)

³ <http://www.uclouvain.be/en-cecl-locness.html>

⁴ BAWE was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics, Warwick), Paul Thompson (Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800), <http://wwwm.coventry.ac.uk/researchnet/BAWE>

⁵ The encoding of the VESPA corpus is largely based on the system developed for the *British Academic Written English* (BAWE) corpus (cf. Ebeling and Heuboeck, 2007).

⁶ The BAWE corpus consists of texts produced by both native (NS) and non-native (NNS) students. We have made use of the NS texts only.

We investigate the frequency and use of features of W/R visibility in these six corpora and focus on first person singular and plural pronouns (*I, me, myself, we, us, ourselves*), second person pronouns (*you, yourself, yourselves*), first and second person possessive determiners (*my, our, your*), *let's* imperatives, epistemic modal adverbs (*certainly, maybe, of course, probably, perhaps*) and questions.

We make use of the Concord tool in WordSmith Tools⁷ to extract features of W/R visibility from the corpora under study. For the annotated corpora (VESPA and BAWE), a tag list was uploaded to WordSmith Tools, so that text within specified tags could be ignored, e.g. text within <quote> and </quote> tags. For ICLE and LOCNESS, which carry no such annotation, some manual sifting of the concordance lines was carried out, in order to make sure that quotes from secondary sources and the like were not part of the material to be analysed.⁸

3. Results

As stated above, we investigate quantitatively some features that have been used as indicators of writer/reader visibility in previous studies and interpreted as signs of an involved writing style. Table 2 shows the frequencies of first and second person pronouns (singular and plural) across the corpora. As Petch-Tyson (1998: 112) found that Scandinavian learners show a higher degree of interpersonal involvement than other learner groups, it was expected that Norwegian learners would have a greater extent of overuse⁹ of personal pronouns than French learners. This proved to be the case for both sets of pronouns in both ICLE and VESPA.

However, there is a significant decrease in the use of first and second person pronouns from argumentative to academic writing for all three L1 backgrounds. This confirmed our hypothesis that assignments in linguistics courses may have invited a less personal writing style than the essay prompts typical of the ICLE corpus.

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE-LING
1 st person	1,898	2,550	791	1,178	1,366	575
2 nd person	344	797	227	41	112	29

Table 2. Personal pronouns across corpora. Frequencies per 100,000 words.

⁷ <http://www.lexically.net/wordsmith/index.html>

⁸ Some manual sifting was also carried out on the BAWE data, as quoted material that is part of the running text has not been marked up in BAWE.

⁹ Whenever the terms 'overuse' and 'underuse' are used, they refer to frequency differences that are statistically significant as shown by a log-likelihood test (p value threshold of 0.01).

Examining the use of individual pronouns, we found that the pronoun *I* did not follow the general trend of decrease in pronoun usage from ICLE to VESPA (cf. Table 3). Both French learners and native speakers use *I* slightly more often in VESPA/BAWE than in ICLE/LOCNESS (though the difference is not statistically significant). The Norwegian learners have a lower frequency of *I* in VESPA than in ICLE, but still use *I* in VESPA with a frequency that is far above VESPA-FR and BAWE-LING.

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE-LING
<i>I</i>	423	767	241	460	703	267
<i>I think</i>	108	108	16	8	18	8

Table 3. ‘*I*’ and ‘*I think*’ across corpora. Frequencies per 100,000 words.

It has been noted by Ringbom (1998: 44), for example, that the expression *I think* is particularly frequent in learner writing and contributes to the high frequency of both *I* and *think*. Table 3 indeed shows that *I think* is responsible for a much larger share of the uses of *I* in ICLE than in VESPA, while the difference between LOCNESS and BAWE is not so great. We can assume that *I think* is used mostly in expressions of opinion, which would explain the clear decrease from ICLE to VESPA (or more generally from argumentative to academic style). Typical examples of this function are given in (1)-(3) below.

1. In my oppinion, *I think* a combination of prison and rehabilitation would be good. (ICLE-NO)
2. All these qualities are of a practical nature and *I think* it is obvious that most university degrees are theoretical and do not prepare students for the real world. (ICLE-FR)
3. *I think* women can serve just as important a role in the military as men can without serving in actual combat duty. (LOCNESS)

Interestingly, Aijmer observes a difference in the function of *I think* between Swedish EFL learners and native speakers: “Learners overuse *I think* in order to make their claims more persuasive and to provide more weight to the issues discussed while native speakers use a less rhetorical style” (Aijmer, 2001: 256). Further explorations of the functions of *I* require more qualitative analysis than we had the opportunity to do for the present study. However, we did take a closer look at the concordances for *I*. More specifically, a search for recurrent word clusters including *I* gave the following results (presented alphabetically; the lists are not exhaustive):

4. *I agree, I am concerned, I am (not) convinced, I am of the opinion, I am sure, I do not think, I mean, I (really, personally) think, I want to, I would say* (ICLE-FR)
5. *I am sure, I believe, I do not/don't think, I feel, I know, I say, I see, I think, I want to, I would like to, I would say, in this essay I* (ICLE-NO)
6. *I am going to analyze/focus/use, I analyzed, I calculated, I can say, I chose to, I collected, I compared, I considered, I consulted, I (have) decided to, I examined, I extracted, I found, I looked, I said, I (have) used, I will analyze/describe/compare* (VESPA-FR)
7. *I believe, I chose, I decided to, I found, I have chosen, I have found, I know, I searched for, I think, I used, I want to, I will (would like to) comment, I will focus on, I will give, I will start, I will use, I would argue, I would like to, I would say, in this essay I will* (VESPA-NO)

The clusters from ICLE, presented in (4) and (5), contain predominantly verbs expressing opinions, as in examples (1) to (3). However, the clusters from VESPA given in (6) and (7) provide a more varied picture: many of the verbs occurring with *I* as subject refer to the procedure of conducting a linguistic analysis and writing up a paper. If this impression is correct, it would imply that the writers present themselves in their texts with different roles or functions depending on the genre of their text. The verbs occurring with *I* thus suggest two main roles that writers take:

- *I* in the role of a senser (Halliday 2004: 201) or arguer, i.e. somebody who offers opinions; this role is more typical of ICLE.
- *I* in the role of a researcher and text organiser; this role is more typical of VESPA.

Some examples of the writer as researcher and text organizer are given in (8)-(11).¹⁰

8. The reason why *I chose* to label these instances as lexical teddy bears ... (VESPA-NO)
9. Every time *I used* Wordsmith to search for a word in the WWC, all the sentences containing the word *I searched for*, seemed to appear three times. (VESPA-NO)
10. Therefore, *I examined* keywords such as ... (VESPA-FR)
11. Finally *I shall analyse and compare* the two corpora (one compiled from academic texts and one from fictional texts) and then draw some conclusions from my analysis. *I shall incorporate* my findings into the theory developed in the first part of my essay.

A comparison between LOCNESS and BAWE yields similar results: there are more occurrences of *I*-clusters used to provide the writer's opinion in LOCNESS than in BAWE, in which *I*-clusters are typically used to describe the different steps of a linguistic analysis, a research procedure or the structuring of the text.

Another feature of W/R visibility which is known to be overused in learner writing (especially among French L1 learners) is the *let's* imperative. As shown in Table 4,

¹⁰ A recent MA thesis (completed after the research for the present study was carried out) lends support to this impression: Fossan (2011: 135) shows that writers come across as predominantly arguers (and next as text organizers and narrators) in ICLE-NO, while writers in VESPA-NO represent themselves predominantly as researchers and text organizers and to a much lesser degree as arguers.

there is a marked overuse of *let's* imperatives in French argumentative essay writing. Paquot (2008; 2010) suggested that this overuse may well be the result of L1 transfer, since imperatives appear to be much more frequent in academic French, where they are used with many different verbs to structure discourse. While there are no statistically significant differences between the Norwegian components of ICLE and VESPA and between LOCNESS and BAWE, there is a significant decrease from 72 to 9 occurrences per 100,000 words of *let's* imperatives produced by French learners in argumentative vs. academic writing. Compared to the BAWE corpus, however, EFL learners' overuse of *let's* imperatives is statistically significant for both French and Norwegian learners but data from the Norwegian component of the VESPA corpus reveals that almost half of the occurrences of *let's* imperatives were produced by the same learner.

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE-LING
<i>Let's</i>	72	15	5	9	9	1

Table 4. '*Let's*'-imperatives. Frequencies per 100,000 words.

Epistemic modal adverbs are a less overt way for writers to appear in their text; however, "in a statement the modality is an expression of the speaker's opinion" (Halliday 2004: 147). Aijmer (2002) found a range of modal expressions, including epistemic adverbs, to be greatly overused by learners as compared to native speakers. Her material was ICLE (particularly ICLE-SW) and LOCNESS. Table 5 shows that, broadly speaking, each L1 background is characterised by a clear decrease in frequency from argumentative to academic (discipline-oriented) writing. This is especially clear from the total frequencies at the bottom of the table. Native speakers make use of fewer epistemic adverbs than learners in both genres and Norwegian learners generally use more epistemic markers than French learners.

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE-LING
<i>certainly</i>	52	24	9	10	4	6
<i>maybe</i>	20	93	16	14	19	2
<i>of course</i>	73	66	13	14	18	5

<i>probably</i>	32	75	19	22	51	9
<i>perhaps</i>	41	34	25	9	39	24
TOTAL	220	293	81	69	130	45

Table 5. *Epistemic modal adverbs. Frequencies per 100,000 words.*

Searching for question marks in the six corpora confirms the patterns of use found for other W/R visibility markers. There are more direct questions in the argumentative than in the more academic-like texts and there are also more direct questions in learner writing than native writing in both genres, which indicates a greater degree of interactiveness in learner writing (cf. Table 6). There is, however, no statistically significant difference between the French and Norwegian texts.

	ICLE-FR	ICLE-NO	LOCNESS	VESPA-FR	VESPA-NO	BAWE-LING
'?	364	387	184	51	55	33

Table 6. *Questions. Frequencies per 100,000 words.*

In summary, there are generally fewer W/R visibility markers in academic or discipline-oriented student texts (that is, in both VESPA and BAWE) than in argumentative essays (such as those found in ICLE and LOCNESS). When compared to their frequency of use in native speaker writing, W/R visibility markers tend to be overused by French and Norwegian EFL learners across genres. However, differences in frequency are often more significant in NNS vs. NS argumentative writing than in NNS vs. NS academic or discipline-oriented writing.

As for differences between French and Norwegian learner writing, we found more personal pronouns and epistemic modal adverbs in Norwegian than in French learner texts. This difference may be explained by differences in writing cultures: as pointed out, for example, by Ramanathan and Atkinson (1999: 55), there is “compelling evidence for widely differing notions of individuality across cultures,” which have “implications for written voice”, i.e. the expression of self in a text. Possibly, Scandinavian writers favour a less formal, more interactive writing style. Another explanation, perhaps, lies in the fact that in both the Norwegian components of the ICLE and VESPA corpora, texts were produced by students who had spent fewer years at university. The Norwegian component of ICLE largely consists of texts produced by 1st-year students while its French counterpart includes texts produced by 2nd, 3rd and even 4th-year students. Similarly, there are fewer MA texts in the Norwegian component of VESPA than in the French sub-corpus and this may be

particularly relevant as MA students in Louvain have an English for Academic Purposes course that is clearly informed by the results of learner corpus research, which means, for example, that students were probably warned against making extensive use of *let's* imperatives.

4. Conclusion

The pilot study presented here is more quantitative than qualitative in nature but already provides interesting insights into EFL learner writing across genres. Norwegian and French EFL learners were shown to use more W/R visibility features and, as a result, to be generally more overtly present within their texts than native students. Our findings thus suggest that the more involved style commonly attributed to EFL learners' writing is not prompted solely by the type of argumentative writing that has usually been analysed in learner corpus research. However, our results also show that learners adapt to genre requirements to some extent, as the frequency of W/R visibility markers was clearly lower in more academic-like (i.e. discipline-specific) writing. As indicated by our brief look at verb phrases used with *I* as subject (section 3), more qualitative studies are needed to survey not only how often learners use features of W/R visibility, but also *how* they use them. The VESPA corpus project will hopefully soon provide us with corpora representing more text types as well as more L1 backgrounds, which will ensure a better basis for studying genre- and L1-specific features of learner writing.

References

- Ädel, A. (2008). Involvement features in writing: do time and interaction trump register awareness? In G. Gilquin, S. Papp and M.B. Díez-Bedmar (eds) *Linking up Contrastive and Learner Corpus Research*. Amsterdam and New York: Rodopi, 35—53.
- Aijmer, K. (2001). *I think* as a marker of discourse style in argumentative Swedish student writing. In K. Aijmer (ed.) *A Wealth of English – Studies in Honour of Göran Kjellmer*. Gothenburg: Acta Universitatis Gothoburgensis, 247-258.
- 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]. Amsterdam and Philadelphia: John Benjamins Publishing Company, 55—76.
- 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*. London and New York: Addison-Wesley Longman, 80—93.

- Ebeling, S. O. and A. Heuboeck (2007). Encoding document information in a corpus of student writing: the British Academic Written English corpus. *Corpora*, 241—256.
- Fossan, H. (2011). “The writer and the reader in texts by Norwegian advanced learners of English. A corpus-based comparison of writer/reader visibility features in Norwegian learner writing and native speaker student writing.” Unpublished MA thesis, University of Oslo. (available at <http://duo.uio.no>)
- Gilquin, G. and M. Paquot (2008). Too chatty: learner academic writing and register variation. *English Text Construction* 1 (1), 41—61.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A.P. Cowie (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press, 145—160.
- Granger, S. and P. Rayson (1998). Automatic lexical profiling of learner texts. In S. Granger (ed.) *Learner English on Computer*. London and New York: Addison Wesley Longman, 119—131.
- 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.
- Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009) *The International Corpus of Learner English*. Version 2. Handbook and CD-Rom. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Halliday, M.A.K. (2004). *An Introduction to Functional Grammar*. Third edition, revised by C.M.I.M. Matthiessen. London: Arnold.
- Hasselgård, H. (2009). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (ed.) *Corpora and Language Teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 121—140.
- Hinkel, E. (2004) *Teaching Academic ESL Writing. Practical Techniques in Vocabulary and Grammar*. Mahwah, New Jersey and London: Lawrence Erlbaum Associates.
- 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., Ballesteros, F., Dafouz, E., Martínez, F. and Rica, J.-P. (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]. Amsterdam: Rodopi, 203—225.
- Paquot, M. (2008). Exemplification in learner writing: a cross-linguistic perspective. In F. Meunier and S. Granger (eds) *Phraseology in Language Learning and Teaching*. Amsterdam: Benjamins, 101—119.
- Paquot, M. (2010). *Academic Vocabulary in Learner Writing. From Extraction to Analysis*. London and New York: Continuum.
- Petch-Tyson, S. (1998). Reader/writer visibility in EFL persuasive writing. In Granger, S. (ed.). *Learner English on Computer*. London and New York: Addison Wesley Longman, 107—118.
- Ramanathan, V. and Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing*, 8, 45–75.
- Recski, L.J. (2004). Expressing standpoints in EFL written discourse. *Revista Virtual de Estudos da Linguagem* 3.
http://www.revel.inf.br/site2007/_pdf/3/artigos/revel_3_expressing_standpoints_in_efl_written_discourse.pdf (last accessed on January 10, 2011).