

The importance of, it is important that or importantly?

The use of morphologically related stance markers in learner and expert writing

Tove Larsson

Uppsala University

The ability to successfully position oneself in relation to one's claims through the use of stance markers is of central importance for academic writers. This study, which uses data from one expert corpus (LOCRA) and three learner corpora (ALEC, VESPA and BATMAT), investigates the use of morphologically related stance markers that occur in different syntactic constructions (such as *possibly*, *the possibility of* and *it is possible that*). In doing so, it examines to what extent lexis, level of expertise in academic writing and L1 transfer influence the distribution of the different realizations of stance under investigation. The results show that all three variables are important predictors. In addition, differences pertaining to information structure are found to influence the distribution of two largely synonymous constructions (disjuncts and the introductory *it* pattern). The findings suggest that there are principled explanations for why one construction is used instead of another functionally similar construction.

Keywords: L2 writing, published expert writing, stance marking, the introductory *it* pattern, disjuncts

1. Introduction

Stance markers function as a means of positioning oneself in relation to one's claims and as such are very important for academic writers to master. Occasionally, however, writers have at their disposal several different realizations of stance that have roughly the same function (e.g. *importantly*, *it is important that* and *the importance of*). Choosing the appropriate realization in a given context can present a challenge for apprentice writers, in particular non-native-speaker writers. The aim of the present study is, firstly, to investigate what factors influence the use of a particular

realization of stance in apprentice learner writing and published expert writing and, secondly, to detect problem areas for learners.

The study uses data from four corpora of linguistics essays and journal articles to explore what influences the distribution of the relevant realizations of stance by investigating variation among stance markers that share the same base form. The apprentice learner data (henceforth ‘learner data’) includes texts from two first-language (L1) groups: Swedish and French. In the first part, the study takes as its starting point five high-frequency base forms (cf. Section 3.2 for a more detailed discussion) from which more than one grammatical realization of stance can be derived. The investigated base forms are *possib** (*possibly, possibility, possible*), *interest** (*interestingly, interest, interesting*), *importan** (*importantly, importance, important*), *presum** (*presumably, presumption, presumed*) and *probab** (*probably, probability, probable*).¹

In the second part of the study, one subtype of the stance complement clause construction and one subtype of the stance adverbials will be placed under scrutiny, namely subject extraposition (henceforth the ‘introductory *it* pattern’; e.g. *it is important that [...]*) and disjuncts (e.g. *importantly*) (cf. Section 4.2.1). These two constructions are the ones that are most similar semantically and syntactically and that therefore enable investigation of what other factors might influence which construction is used in a given context. For this part of the study, the base forms *clear** and *surprising** will replace *presum** and *possib**, as the latter pair exhibited frequencies for the introductory *it* pattern that were too low to be investigated (cf. Section 4.2.1).

2. Previous research on stance marking

The present study uses the term ‘stance’ (as realized by stance markers) to denote the means by which writers convey “a textual ‘voice’” to express commitment to or assessment of a proposition (Hyland 2005: 176). The study considers three of Biber et al.’s (1999: 969–970) five main categories of stance marking constructions (henceforth ‘grammatical realizations of stance’): (i) stance adverbials (e.g. *unfortunately*),

1. While the present study focuses on only the five base forms that were common to both the learner and the expert corpus (see Section 3.2) in order to look at the factors outlined above, these base forms are by no means the only ones of this sort. In fact, morphologically related stance markers seem to be a pervasive phenomenon; searches in a considerably larger, 520-million-word corpus, Corpus of Contemporary American English (COCA), yielded a vast array of such triplets, including *predictab** (*predictably, predictable, predictability*), *certain** (*certainly, certain, certainty*), *inevitab** (*inevitably, inevitable, inevitability*), *curious** (*curiously, curious, curiousness*) and *convenien** (*conveniently, convenient, convenience*).

(ii) stance noun + prepositional phrase (e.g. *they deny **the possibility** of increased sales [...]*) and (iii) stance complement clauses controlled by a verb (e.g. *I **hope** that [...]*), a noun (e.g. ***the fact** that [...]*) or an adjective (e.g. *I'm **happy** that [...]*); the third category includes extraposed structures (e.g. *it is **interesting** that [...]*). For some base forms of stance markers (i.e. a lexical root to which derivational and/or inflectional morphemes can be added), writers can use any of several different grammatical realizations of stance, each belonging to a different category in Biber et al.'s (1999) framework. The three grammatical realizations of stance are exemplified for the base form *possib** in the constructed sentences in Examples (1) to (3).²

- (1) You could *possibly* go there late in the evening.
- (2) You could have *the possibility* of going there late in the evening.
- (3) It is *possible* that you could go there late in the evening.

This raises the question of why one is used instead of another. This question has received very limited attention in the literature.

Several studies have, nonetheless, investigated the use of stance markers in more general terms. Evaluative language use has been investigated in several different paradigms, including stance (e.g. Biber et al. 1999, Biber 2006b, Hyland 2005), attitude (e.g. Halliday 1994), appraisal (e.g. Martin & White 2005), and evaluation (e.g. Hunston & Thompson 2000). Furthermore, the object of study also varies. While some studies focus on one subgroup of stance marking, such as nouns preceded by a sentence-initial deictic (e.g. *This **technique***) (Charles 2003), others take a more inclusive approach. Examples of the latter include the studies carried out by Biber (2006a, 2006b) and Biber et al. (1999:969), where grammatically signaled stance marking is investigated. Grammatical stance marking denotes “a stance relative to some other proposition” (Biber et al. 1999:968), as in *I am **happy** that he came in early*. This kind of stance marking thus differs from affective words, such as *grateful* in *I am **grateful***, as the function of such words is “to directly attribute emotional or attitudinal stance to the speaker” (Biber et al. 1999:968). Grammatical stance marking is the framework that will be used in the present study, as described in Section 3.2.

With regard to linguistic and extralinguistic variables that can influence the grammatical realization of stance, three main factors of potential importance emerge from previous literature: lexis (as realized here by the base forms), level of expertise in academic writing (expert vs. apprentice writers) and negative L1 transfer (i.e. language interference from the learner's first language in his/her use

2. Italics and/or bold face has been added to certain examples throughout the article to point the reader's attention to the relevant sentence element(s).

of another language). First, since lexis and grammar are often found to be inseparable (e.g. Sinclair 1991, Hunston & Francis 2000), it is not unlikely that lexis will be a factor of importance in this study as well; that is, the five base forms are likely to differ with regard to which grammatical realization is more frequent. Second, differences between expert and apprentice writing with regard to different kinds of stance markers have been noted in, for example, Neff et al. (2003), where advanced EFL learners with various L1s were found to overuse *it is* + ADJ + *that* (e.g. *it is true that* [...]) and conjuncts, such as *however* and *nevertheless*. Third, certain signs of L1 transfer have been found in the use of adverbials in L1 Norwegian student writing (Hasselgård 2009) and the stance complement clause construction in L1 Greek student writing (Hatzitheodorou & Mattheoudakis 2009). Adverbials and the stance complement clause construction are investigated in the present study too; there is therefore reason to believe that L1 transfer may play an important role here as well.

In the light of the findings of previous studies, the following research questions will be addressed in this study:

- i. To what extent do lexis, expertise in academic writing and L1 transfer influence the distribution of the grammatical realizations of stance?
- ii. In addition to these three factors, what other factors may affect the distribution of two semantically and syntactically very similar constructions, namely the introductory *it* pattern and disjuncts?

3. Corpora and method

In Section 3.1, the four corpora used for the present study will be introduced, while in Section 3.2 the method will be described in detail.

3.1 Corpora

The corpora used for the present study include texts from only one academic discipline, namely linguistics. Although this approach may limit the generalizability of the study, it makes it possible to rule out disciplinary differences as a potential explanatory factor for any differences found. Controlling for disciplinary differences is clearly advantageous, as important differences across disciplines have been noted with regard to the use of stance markers in general (e.g. Hyland 2005), as well as for certain constructions that can be used for stance marking (e.g. Larsson under review, Groom 2005).

The corpora comprise journal articles and student essays written by published expert writers and apprentice learners of English respectively. The texts were

sampled from four corpora: the Louvain Corpus of Research Articles (LOCRA), the Advanced Learner English Corpus (ALEC), the Varieties of English for Specific Purposes dAtabase (VESPA) and the BATMAT corpus. LOCRA is a corpus of journal articles in business, medicine and linguistics that is under compilation at the Centre for English Corpus Linguistics at Université catholique de Louvain in Belgium (<http://www.uclouvain.be>). Only articles that were published in high-prestige journals (i.e. journals that were highly regarded by researchers in the respective fields at Université catholique de Louvain) were included; the linguistics subcorpus used in the present study comprises articles from a variety of different subfields of linguistics, such as applied linguistics and sociolinguistics. The material is not restricted to native-speaker writing. ALEC comprises learner writing from university students in English linguistics and literature in Sweden (Larsson 2014). VESPA is an international corpus of learner writing administered at the Centre for English Corpus Linguistics at Université catholique de Louvain in Belgium. The corpus is currently being compiled and includes texts from several European countries, for example Belgium, Sweden and Norway (Paquot et al. 2013). The BATMAT corpus is also still being developed at Åbo Akademi University in Finland and comprises texts written by students in English linguistics, English literature and realia (Lindgrén 2015).

From these corpora, two main subcorpora (consisting solely of linguistics texts) were compiled: one with expert data from LOCRA, and the other with learner data from the three remaining corpora. The average length of the texts is 8,200 words for the experts and 8,500 for the learners. The vast majority of the texts included in the learner corpora (100 of 114, or 88%) were written by students who were in their third year of university study. In order to investigate possible transfer-induced problems, the learner data was further subdivided into two subcorpora: one with L1 Swedish learner data from ALEC, BATMAT and VESPA, and the other with L1 French learner data from VESPA.

Although there were certain minor differences between the L1 Swedish corpora, such as the BATMAT students' lesser tendency to use stance noun + prepositional phrase compared to the other L1 Swedish corpora, the differences were not statistically significant ($p > 0.05$; Pearson's Chi-squared test); all three L1 Swedish subcorpora were therefore merged into one L1 Swedish corpus. It should also be noted that the L1 French subcorpus is small by comparison; this will be returned to when the results are discussed. Both learner subcorpora were sampled to ensure comparability as much as possible with regard to length, text type and level (i.e. year of study). This careful sampling did, however, result in a reduction in data. An overview of the subcorpora is provided in Table 1.

Table 1. Subcorpora included in the study

Subcorpus	Level of expertise in academic writing	L1	Number of texts	Number of words
Expert	Expert	N/A	109	997,557
L1 Swedish	Apprentice	Swedish	94	772,917
L1 French	Apprentice	French	20	103,840
Total			223	1,874,314

3.2 Method

In order to find the morphologically related stance markers that can occur in different grammatical realizations of stance, the study needs to establish (i) what grammatical realizations of stance there are and (ii) which base forms can occur in all of these grammatical realizations. For (i), Biber et al.'s (1999:969) framework of grammatical stance marking is used. Of the five categories in this framework, the study investigates members from the three categories for which there is potential overlap with regard to morphologically related stance markers: stance adverbials (e.g. *interestingly*, *possibly*), the stance complement clause construction (e.g. *I am happy that, it is important to [...]*) and stance noun + prepositional phrase (e.g. *the importance of [...]*).³ In order to identify which base forms can occur in all three of these grammatical realizations, i.e. to address (ii), a bottom-up approach was applied, as described below.

Since the corpora are not part-of-speech (POS) tagged, search expressions were used. The corpora used did not come with any corpus programs to access them; instead, the concordancing program AntConc (Anthony 2014) was used, along with scripts written in the computer programming languages *R* and *Perl*. The stance complement clause category encompasses clauses that are controlled by verbs (e.g. *I hope that [...]*), adjectives (e.g. *we are sad that [...]*) and nouns (e.g. *the fact that [...]*); this category also includes instances of the introductory *it* pattern (e.g. *it is important to [...]*). The following search expression was used to find occurrences of this category of grammatical realizations of stance: `[I|we|the|it] [\w]{1,3} [to|that]` (i.e. *I, we, the* or *it*, followed by one to three words, followed by *to* or *that*). Following Biber (2006a: 100), *wh*-clauses and participial clauses have been excluded. From the stance adverbial category, the subcategory that has been found to make up the vast majority of the tokens in this category (Biber et al. 1999:982), single adverbs,

3. The study does not, thus, include investigation of the two remaining categories: modals and semi-modals (e.g. *might*; *has to*) and premodifying stance adverb (stance adverb + adjective or noun phrase) (e.g. *I'm so happy for you*; *fossils of about this age have been found*).

was selected for analysis. For this subcategory, all words ending in *-ly* have been searched for. The final category, stance noun + prepositional phrase, was searched for using the following search pattern: the $[\backslash w]\{1,3\}$ of (i.e. *the* followed by one to three words, followed by *of*).

There are of course limitations of using search patterns. For instance, the ones used here do not capture stance markers with many intervening lexical items, as in Example (4), nor do the search patterns capture multi-word adverbials (e.g. *in actual fact*). However, such instances are much less frequent than the instances that are captured by the search patterns.

- (4) It is, at the very least, not conceptually necessary, in acknowledging a relation between [a] and [β], to reify (objectivize) that relation as a third object [...].
(LOCRA_008-03)

Furthermore, even though the use of search expressions results in a more limited scope, each search expression returned between 18,000 and 31,000 hits, resulting in a total of approximately 80,000 potentially valid stance markers. In order to process the data, the software environment *R* (R Core Team 2015) was used. Using the first four letters of each word returned by the search expressions, *R* was utilized to take the intersection of the hits, i.e. to summarize only the hits that were common to both the expert and the learner subcorpora and that were found to be instantiated by all three grammatical realizations of stance (i.e. complement clause, adverbial and noun + prepositional phrase) (see Section 4.2 for the slightly adapted approach used for the second part of the study). This step resulted in a list of 138 such four-letter types (e.g. *cons**, *comm** and *part**). The list was subsequently checked manually to decode the four-letter hits, as, for example, *cons** covers no fewer than 16 different lexical types, ranging from *constituency*, *constancy* and *construction* to *constraint* and *consecutively*. Manual scrutiny of the results was necessary, as only a very limited number of the lexical items found were stance markers, were common to both subcorpora and were common to all three grammatical realizations of stance.

Stance is a broadly used concept, and the definition of what constitutes a stance marker varies across studies. In order to make this study more easily replicable, a decision was made to use Biber's (2006b: 92–93) extensive list of grammatical stance markers. The list of potential stance markers in the present study was checked against Biber's (2006b) list. Triplets (such as *simplistically*, *simplification*, *simplicity*) that did not contain at least one member that was included on this list and that were thus not perceived as carrying stance were removed. Once all these steps had been carried out, five triplets of stance markers remained: *interest** (*interestingly*, *interesting*, *interest*), *importan** (*importantly*, *important*, *importance*), *possib** (*possibly*, *possible*, *possibility*), *presum** (*presumably*, *presumed*, *presumption*) and *probab**

(*probably, probable, probability*). The second part of the study investigated a slightly different set of stance markers; see further Section 4.2.1.

The differences found were subsequently tested for statistical significance using *R*. Since most of the variables examined are nominal (i.e. categorical rather than numeric), tests suitable for these kinds of data were chosen. Pearson's Chi-squared test was used to test for overall differences and the Kruskal-Wallis rank sum test was used to test differences between medians. Two models were also fitted: one binomial generalized linear model to test how the independent variables (e.g. lexis, L1) and their interactions affect a dependent variable with two levels (e.g. type of construction), and one multinomial log-linear model to test how the independent variables (and their interactions) affect a dependent (nominal) variable with more than two levels (e.g. type of grammatical realization of stance). Finally, searches were carried out in the Europarl parallel corpus (Koehn 2005) to double-check the translations of the investigated constructions discussed in the present article.

4. Results and discussion

The distribution of stance markers across all three grammatical realizations of stance (i.e. the complement clause construction, adverbial and noun + prepositional phrase) will be investigated in Section 4.1, after which the introductory *it* pattern and disjuncts will be placed under further scrutiny in Section 4.2.

4.1 Stance marking across the three grammatical realizations

Five base forms of stance markers are included in this part of the study (cf. Section 3.2): *interest** (*interestingly, interesting, interest*), *importan** (*importantly, important, importance*), *possib** (*possibly, possible, possibility*), *presum** (*presumably, presumed, presumption*) and *probab** (*probably, probable, probability*). Each of these base forms can be instantiated through all three relevant grammatical realizations of stance: stance adverbials, stance noun + prepositional phrase and the stance complement clause construction. There were 1,821 tokens found in the data.

In Section 4.1.1, the results for the first variable, lexis, will be examined. In Section 4.1.2, the results for the remaining two variables, level of expertise in academic writing (expert vs. apprentice writing) and L1 transfer (L1 French vs. L1 Swedish learners of English) will be discussed together, as these two variables are closely related.

4.1.1 Lexical influence

When the frequencies for each base form are investigated, clear differences can be noted in the overall frequencies, as well as in the relative frequencies of the three grammatical realizations of stance. Figure 1 displays the raw frequencies across the five base forms in all the corpora. In the tables and figures, “ly” will henceforth be used to denote stance adverbials; “the of” will be used for stance noun + prepositional phrase; and “comp” will be used for the stance complement clause construction. Moreover, the first four letters of each base form are used in the figures (rather than the full base form) for ease of reading.

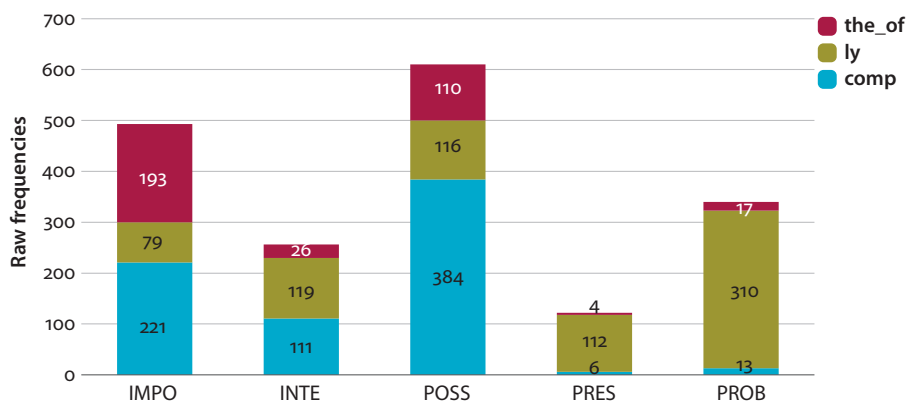


Figure 1. Raw frequencies for each of the grammatical realizations of stance across the five base forms

As can be seen, *presum** and *probab** are almost exclusively realized as stance adverbials, whereas *possib** is mainly realized as a stance complement clause construction. The base form *importan** stands out as the only one that is very commonly realized through the stance noun + prepositional phrase construction, although the most common realization is still the stance complement clause construction. The base form *interest** is almost equally frequently realized through stance adverbials and the stance complement clause construction, although stance adverbials are slightly more common. The pairings of preferred realizations for each base form are shown in Examples (5) to (9) below.

- (5) Much of the practice *probably* occurred incidentally without conscious attention [...]. (LOCRA_013-04)
- (6) *Presumably*, this is also true for L2 readers [...]. (ALEC_085)
- (7) [...] *it is entirely possible that* the inferential possibilities are different in the case of independent sentences [...]. (LOCRA_001-03)

- (8) [...] *it is important to have* a varied language [...]. (VESPA.Swe_0137-01)
- (9) *Interestingly*, the twins differed with respect to rate and pattern of acquisition [...]. (LOCRA_001-02)

The fact that there are such clear inter-lexical differences suggests that lexis is a factor of importance to take into consideration when investigating constructions or syntactic categories.

In addition to frequency differences, further inter-lexical variation is found when the results are investigated in more detail. This is especially clear for the stance adverbials, for which two main differences are evident: one pertaining to common usage patterns and the other pertaining to placement. For example, whereas *interestingly* is rarely modified, *importantly* frequently is, very commonly by *more* or *most*, as in Example (10). *Possibly* is frequently preceded by a modal verb, as in Example (11). *Presumably* and *probably* are commonly found in-between a linking verb and a subject predicative in a clause (S+V_{linking} + {*presumably, probably*} + sP), as in Examples (12) and (13).

- (10) More *importantly*, employees will be assumed to possess such knowledge [...]. (LOCRA.016-02)
- (11) This could *possibly* explain why the layer [...] is 30 times more frequent [...]. (LOCRA.007-05)
- (12) This phenomenon is *presumably* orthogonal [...]. (LOCRA.002-02)
- (13) English is *probably* one of the world's most spoken languages [...]. (ALEC_009)

With regard to sentence placement, some further differences can be noted. While *importantly* and, in particular, *interestingly* are very often placed sentence initially (i.e. without any preceding lexical items), *possibly*, *presumably* and *probably* are only very rarely sentence-initial. Of the few instances of sentence-initial *possibly*, *probably* and *presumably*, the L1 Swedish learners were responsible for almost all the occurrences (no occurrences were found in the L1 French data); many of these instances were furthermore marked or not grammatically correct. Since the Swedish equivalents (e.g. *möjligtvis*, *troligen* and *förmodligen*) can be placed sentence-initially, the results suggest that L1 transfer is involved, which is something that will be discussed further in the next subsection. An example of sentence-initial use of *troligen* from the Europarl corpus can be found in Example (14), with a translation in Example (15).

- (14) *Troligen* är det en av de saker som vi måste gottskriva den politik som Mugabes regering förde under sina första år. (Europarl_sv-en.sv)

- (15) This is *probably* one of the things we should give the Mugabe government credit for during its early years. (Europarl_sv-en.en)

Whether the preferences discussed above extend to both expert and learner writing more generally will also be examined.

4.1.2 Level of expertise in academic writing and L1 transfer

Indeed, as shown in Figure 2, preferences differ between expert and learner writing. Overall, the experts made use of the stance markers investigated 770 times (772 pmw), and the learners used them 1,051 times (1,199 pmw).

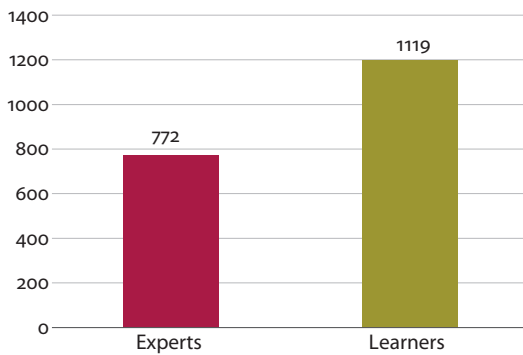


Figure 2. Overall frequencies of the investigated stance markers per million words for the experts and the learners

A closer look at the results shows that the differences extend to the level of the individual too, as can be seen in Figure 3, where the per-text frequencies of the investigated stance markers are shown in notched box plots. Each box represents the fifty percent of the data closest to the median (marked as a black line); the notches mark the confidence interval around the median. The whiskers (the vertical bars extending from the boxes) mark the minimum/maximum (or 1.5 times the inter-quartile range below/above the box), and the dots represent outliers.

The results show that compared to the experts, the learners make significantly more frequent use of stance marking (as realized through the categories investigated) both overall⁴ and per text.⁵ These findings are in line with those of previous studies looking at learner use of stance marking in more general terms, albeit in comparison to native-speaker apprentice writing, where the learners were

4. Pearson's Chi-squared test; $p < 0.001$, Chi-squared: 87.595.

5. Kruskal-Wallis rank sum test; $p < 0.001$, Chi-squared: 16.678.

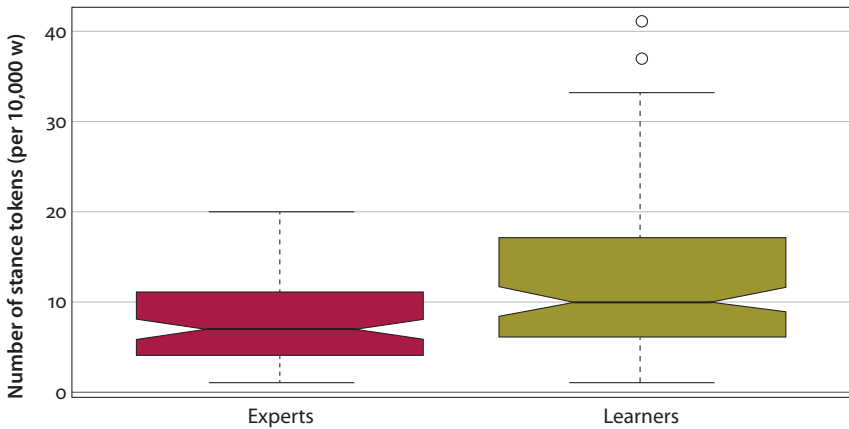


Figure 3. Normalized frequency (per 10,000 words) of all stance tokens per text in expert writing (left) and learner writing (right)

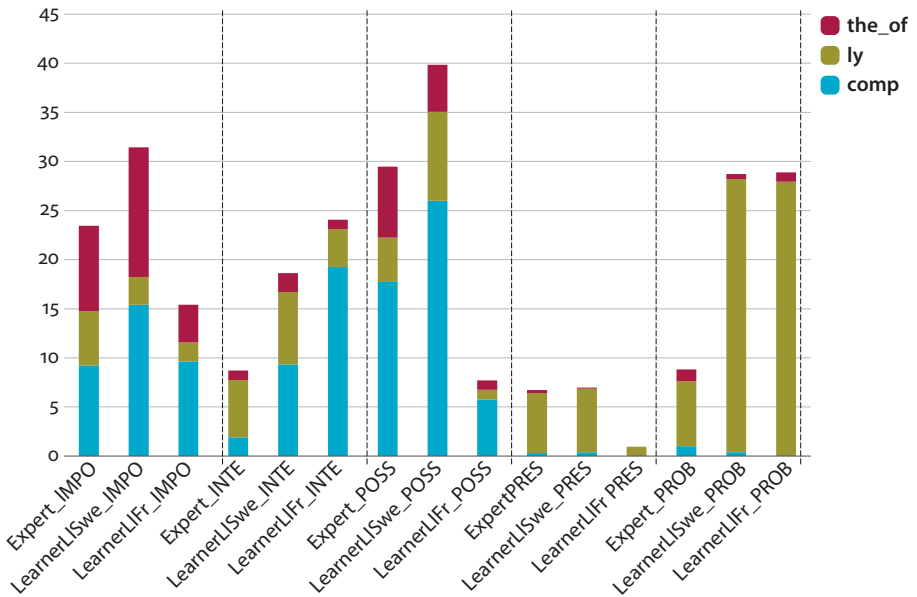


Figure 4. Frequencies per 100,000 words for each of the grammatical realizations of stance across the five base forms for the three groups

reported to encode more interpersonal information in their texts (Hasselgård 2009, Herriman & Boström Aronsson 2009).

Additional notable differences can be found when the L1 Swedish group is separated from the L1 French group for each of the grammatical realizations of stance

across the five base forms, as can be seen in Figure 4. As mentioned in Section 3.1, the L1 French subcorpus is considerably smaller than the L1 Swedish subcorpus. While the frequencies shown in the figure are normalized, it should be kept in mind that the L1 French data comes from a smaller sample; the raw frequencies can be found in Appendix 1 for comparison.

In particular, there are statistically significant⁶ differences in overall frequency between the experts and both learner groups for *interest** and *probab** and between the experts and the L1 French students for *possib**. Furthermore, instances of *presum** were almost non-existent in the L1 French subcorpus; only one occurrence was found in the data. The use of the base forms *interest** and *probab** will be discussed in more detail below. The base forms *possib** and *presum** will also be addressed briefly.

With regard to *interest**, the learners do not only overuse the base form on the whole, but also use the grammatical realizations of stance differently. While the most frequent realization for the experts is the stance adverbial, the most frequent one for the learners is the complement clause construction, in particular the introductory *it* pattern, as shown in Examples (16) and (17) respectively.

(16) *Interestingly*, Tables 15 and 16 reveal that the D values for the learners in London are higher on all tasks [...]. (LOCRA_022-04)

(17) *It is interesting to see* that the Scandinavian languages could have influenced the English usage of modal auxiliaries in Middle English. (VESPA.Swe_0018 01)

Furthermore, the learners' general overuse of *interest** can largely be explained by the fact that the learners use this base form several times per text. In fact, no less than 39 percent (45/114) of the students used *interest** twice or more in their text, as compared to 20 percent (22/109) for the experts.

The normalized frequencies for *interest** per text across the three grammatical realizations of stance in the learner and expert writing are shown in Figure 5. The differences between the groups are statistically significant.⁷

In addition, while the experts mainly use one stance category (typically the stance adverbial form) of *interest**, the learners very frequently use members of two or even three of the categories in the same text.

There is some evidence that the learners' tendency to use the stance adverbial form for *interest** comparably infrequently could be due to L1 transfer. There is no

6. Multinomial log-linear model; model AIC: 5389.466; the coefficients and standard errors of the model can be found in Appendix 2.

7. Kruskal-Wallis rank sum test; $p < 0.01$, Chi-squared: 9.2567.

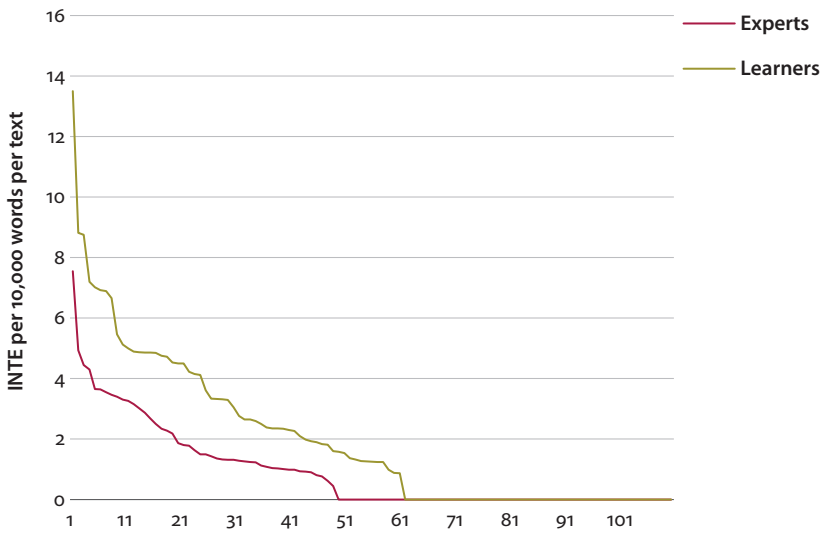


Figure 5. Frequencies per 10,000 words of the base form *interest** per text in the learner and expert data

direct and idiomatic equivalent of *interestingly* in Swedish or French; instead, one would have to say *intressant nog* (“interesting(ly) enough”) and *fait intéressant* or *assez intéressant* (“interesting fact” or “interesting enough”). Instances from Europarl can be found in Examples (18) to (21).

- (18) Vi fick *intressant nog* veta att omstruktureringar [...] kunde genomföras mer effektivt [...].
(Europarl_sv-en.sv_010517)
- (19) We found out, *interestingly*, that restructuring [...] can be implemented more efficiently [...].
(Europarl_sv-en.en_010517)
- (20) *Fait intéressant*, la population irlandaise s’élevait à 8 millions d’individus en 1845.
(Europarl_fr-en.fr_080220-012)
- (21) *Interestingly*, in 1845 the Irish population was over 8 million.
(Europarl_fr-en.en_080220-012)

Therefore, the learners can be expected to do one or both of the following, if L1 transfer is involved: first, when wishing to express the semantics of *interest**, the learners are more likely to choose the construction they have in their L1 (i.e. the stance complement clause construction) than one they do not have, which appears to have been the case, as both learner groups made more frequent use of the stance complement clause construction than the stance adverbial form of *interest** compared to the experts. Second, if they do use the construction that does not have

a verbatim correspondence in their L1, they are likely to use that construction in a way that is similar to the way it is used in their L1; we could then expect both learner groups to make use of the pair *interestingly enough* to a larger extent than the experts. The L1 French learners could also be expected to use *interesting fact*.

Although there are only relatively few occurrences, a closer look at the data seems to support this claim, as approximately 60 percent (10/17) of the instances of *interestingly* in the L1 Swedish data and both (2/2) instances in the L1 French data were followed by *enough*, as in Example (22). This can be compared to only one instance out of 58 of *interestingly* followed by *enough* in the expert data. Moreover, a separate search for *interesting fact* in the L1 French data shows that there are a few instances of this word pair, as shown in Example (23). The learner groups' use of *interest** thus seems to have been affected by L1 transfer.

- (22) *Interestingly enough*, the example above allows for both these interpretations
[...]. (ALEC.3.033)
- (23) However, *the interesting fact* is that students alternate both constructions in their writing spontaneously. (VESPA.Fr_1305-01)

When it comes to *probab**, the learners make significantly more frequent use of the stance adverbial form (*probably*) than the experts.⁸ The very high frequencies of *probably* in the learner data proved to be due to a handful of students' very frequent use of this form. One example of a learner's use of *probably* is given in Example (24). Two possible explanations for the difference in use of *probab** between the learners and experts will be discussed below.

- (24) [...] we do not know exactly how much, but the amount of love required is *probably* enormous. (VESPA.Fr_1301-01)

First, it could be due to varying levels of expertise in academic writing. The reason for the experts' infrequent use of *probab** could be that this base form is associated with spoken discourse, and the experts might therefore consider it too informal for academic writing. A search in large general corpus of American English, the Corpus of Contemporary American English, COCA, 520 million words as of 2016 (Davies 2008), seems to confirm this assertion, as *probably* is strongly associated with the spoken genre (*probably* occurs 397 times per million words in the spoken corpora, compared to 133 times per million words in academic texts). This concurs with Conrad & Biber's (2000: 64) finding that *probably* is more strongly associated with conversation than with academic prose and news. It can also be noted that the adverbial forms of the other base forms are either more strongly associated

8. Pearson's Chi-squared test; $p < 0.001$, Chi-squared: 126.99.

with academic discourse than spoken discourse (this is the case for *interestingly*, *importantly* and *presumably*) or only somewhat more frequent in spoken data (as is the case for *possibly*) in COCA. A clear overuse of *probably* has also been reported in other studies comparing NNS writing to NS writing, such as in Hasselgård's (2009: 135) study of L1 Norwegian learners; it was also shown in this study that *probably* exhibits high frequencies in spoken NS data. It might then be the case that the learners whose texts are included in the present study may have somewhat lacking register awareness, which has been noted also in Herriman & Boström Aronsson (2009: 118). They even go as far as to state that the L1 Swedish students' argumentative writing "is similar to NS' spoken language" (Herriman & Boström Aronsson 2009: 118).

Second, the differences in the use of *probab** between the learners and experts might also be due to L1 transfer, in particular for the L1 French group. As both *presumably* and *probably* commonly translate as *probablement* in French, this group of learners could be expected to choose the version that is more similar to the French equivalent, namely *probably*, instead of *presumably*. If this is the case, this might explain not only why the L1 French learners make much more frequent use of *probab** than the experts, but also why the L1 French learners make very infrequent use of *presum**.

Further evidence for L1 transfer being an important variable can be found at a more detailed level when we look at the L1 Swedish students' use of *possib**. Negative transfer might be what lies behind the L1 Swedish learners' greater tendency to put the adverbial form of *possib** (and, to a certain extent, *presum** and *probab**) in sentence-initial position, as was mentioned in Section 4.1.1. In the learner data, the L1 Swedish learners were responsible for all of these instances of *possibly*, *presumably* and *probably*; there were no instances found in the L1 French data. An instance of sentence-initial *possibly* from the L1 Swedish data is given in Example (25).

- (25) *Possibly* it means that many teachers seem to have embraced the idea that that "good teachers" do not follow textbooks [...]. (ALEC_013)

Common translations for *possibly* in Swedish include *möjligtvis* and *möjligen*, as in Examples (26) and (27) from Europarl.

- (26) Dessutom skulle grupper av småföretagare *möjligen* kunna samarbeta och registrera tillsammans. (Europarl_sv-en.sv_000314)
- (27) In addition, groups of similar small business could *possibly* cooperate and register together. (Europarl_sv-en.en_000314)

Since *möjligtvis* and *möjligen* can be placed sentence-initially in Swedish, the tendency to sometimes place *possibly* sentence-initially might be explained by L1 transfer.

In sum, the analyses reported here suggest that all three variables that have been investigated so far help to explain the distribution of the three grammatical realizations of stance. However, as the discussion shows, the variables are, at times, difficult to tease apart. In order to further examine what could influence the distribution of stance markers, the study will now zoom in on the two most similar constructions: the introductory *it* pattern and disjuncts. Doing so allows us to investigate the second research question, namely what other factors, in addition to those investigated above, may affect the use of these two stance-marking constructions.

4.2 The introductory *it* pattern vs. disjuncts

In Section 4.2.1, the two constructions – the introductory *it* pattern and disjuncts – are introduced and the question of what factors (other than the ones discussed in 4.1) might affect their distribution is addressed. As will be further explained below, these factors were subsequently removed after classification in preparation for Section 4.2.2, where we return to and test the impact of the predictors discussed in Section 4.1.

4.2.1 *Establishing a basis for comparison*

The introductory *it* pattern and disjuncts have been described as being very similar constructions both syntactically and semantically. In fact, they have even been described as being paraphrases of one another (Quirk et al. 1985: 614, 623). Both constructions have a superior role in relation to the clause they are commenting on (Quirk et al. 1985: 614). The two constructions can then be expected to be similar enough to be in competition with each other, while still exhibiting characteristics that make them slightly different, therefore making one construction better suited than the other in certain contexts. What factors might be important for predicting the distribution of the constructions will be investigated in the present subsection.

In Quirk et al.'s (1985: 501) terminology, disjuncts, such as *interestingly* in Example (28), make up one of four subcategories of grammatical functions of adverbials; the remaining categories are conjuncts, subjuncts and adjuncts. Disjuncts have a superior role in relation to the other sentence elements, in the sense that their scope can be considered to span the whole sentence (Quirk et al. 1985: 613).

- (28) *Interestingly*, the patterns of subject and verb tense were very uniform in TQ
[...]. (LOCRA_017-05)

The introductory *it* pattern, similarly, takes a superior role in relation to the other sentence elements, as shown in Example (29).

- (29) [...] *it is interesting that* both the sentences above display disagreement [...].
(ALEC_083)

This pattern has also been referred to as ‘subject extraposition’ (e.g. Quirk et al. 1985: 1391, Biber et al. 1999: 155). The pattern is here defined as being made up of two subjects: one introductory *it* that does not have anaphoric reference, and a nominal clause (see Larsson 2016 and Larsson forthcoming, for a more detailed discussion). In terms of function, this pattern, especially when it includes an adjective, has been found to be strongly associated with evaluation (Hunston & Sinclair 2000: 84).

Since the base forms *presum** and *probab** proved to be strongly biased towards the stance adverbial form, there were too few instances of the introductory *it* pattern to include these base forms in this part of the investigation. Instead, two other base forms, *clear** (*clearly* and *it is clear that*) and *surprising** (*surprisingly* and *it is surprising that*), were added to the remaining three base forms (*important**, *interest** and *possib**) to increase the number of base forms investigated. There were 1,305 valid tokens of these five base forms found in the data. The reason why *clear** and *surprising** are only included in the second part of the study is that they did not meet the criteria for inclusion in the first part of the study (by not occurring frequently enough in the stance noun + prepositional phrase construction); they are the only remaining adjective-adjective pairings to meet the criteria for inclusion in the second part of the study.

Through manual, qualitative investigation of the data, four factors stood out as particularly important for predicting which of the two constructions is used: the possibility of (i) adding additional information through a *to*-infinitive clause, (ii) marking tense, (iii) using certain modifiers, and (iv) using pre-clausal hedges. These factors are exemplified in Exmples (30) to (33) below.

- (30) *It is interesting to hear* women say these things. (ALEC.3.073)
- (31) [...] *it was clear that* the term originally stemmed from English.
(VESPA.Swe_0114-01)
- (32) *It is entirely possible that* the inferential possibilities are different.
(LOCRA.001-03)
- (33) But *it seems clear that* the affricate involves a greater amount of movement
[...]. (LOCRA.015-03)

Since these four factors lead writers to use only one of the constructions, namely the introductory *it* pattern, they were considered knockout factors. In the next

subsection (Section 4.2.2), the tokens that contained any of these (with the exception of tense, as discussed in Section 4.2.2) were therefore removed after classification, to investigate the potential influence of the predictors discussed in the previous section (inter-lexical variability, level of expertise in academic writing and the students' L1) with greater precision. First, however, the four factors will be discussed briefly in turn.

The first factor to be addressed is the possibility of adding information through an infinitival verb, as in Example (30) above. In the data, there were three main categories of instances of the introductory *it* pattern that include a *to*-infinitive clause: those with a subject, as in Example (34); those without a subject, as in Example (36); and those without a subject, but with an embedded clause immediately following the infinitival verb, as in Example (38).

However, for all three subtypes, the semantics of the introductory *it* pattern with a *to*-infinitive clause is arguably considerably different from that of the equivalent disjunct, as in the corresponding constructed Examples in (35), (37) and (39). This suggests that the two constructions are not in competition; adding information through a *to*-infinitive could therefore be considered a knockout factor.

(34) [...] *it is important for them to remember* how much they and their followers have fought [...]. (ALEC.3.075)

(35) *Importantly*, they remember how much they and their followers have fought.

(36) [...] *it is interesting to look* at other factors in 'personal identity' [...]. (ALEC.4.003)

(37) *Interestingly*, other factors in 'personal identity' are worth looking at.

(38) *It is interesting to note* that this standpoint was taken as early as in 1994 [...]. (ALEC.4.013)

(39) *Interestingly*, this standpoint was taken as early as in 1994.

With regard to the second factor, added modifiers, there were certain differences in the data between the constructions that could explain their distribution. Whereas both constructions were found to be modified by adverbs such as *more* and *most*, the majority of the adverbs in the predicate of an introductory *it* pattern were not used to modify disjuncts. Examples include adverbs such as *certainly*, *entirely*, *especially*, *hardly*, *generally* and *immediately*, two of which are shown in Examples (40) and (41). Thus, if writers wish to convey the semantics of one of these modifiers in combination with the investigated base forms, the introductory *it* pattern seems to be the best alternative.

(40) *It is immediately* clear that such a lexical item will be able to combine with a range of other items [...]. (LOCRA_006-03)

- (41) It is *hardly* surprising that the international student prospectus imitates the rhetoric of advertising [...]. (LOCRA_019-05)

Similarly, with regard to the third factor, marking tense, the introductory *it* pattern seems to be the easiest alternative, as the pattern, unlike disjuncts, allows writers to mark tense inside the construction. The vast majority of the instances of the introductory *it* pattern in the data were found to be in the present tense; only 91 out of 796 tokens were in the past tense. The past tense is therefore considered the marked alternative. Two instances where the past tense is used are provided in Examples (42) and (43).

- (42) It *was* also important that the interview questions were [...] valid [...]. (ALEC_066)
- (43) [...] it *seemed* clear that OQF 'tomship' came from English 'township' [...]. (LOCRA_012-03)

There are also cases where the tense differs between the matrix clause and the extra-posed clause as in Example (38) above (*It is interesting to note that this standpoint was taken as early as in 1994 [...]*), which is something that is only possible for the introductory *it* pattern. However, as this use was found to be very infrequent (only a couple of tokens were found in the data – and then in conjunction with one of the other knockout factors), it will not be discussed further here.

The token displayed in Example (43) can also serve to show the final factor: pre-clausal hedges. Some of the instances of the introductory *it* pattern in the data were found to include a pre-clausal hedge, such as *SEEM* and *APPEAR*. As such hedges were not found to occur in conjunction with disjuncts, the two constructions do not appear to be in competition for these kinds of tokens. Modal auxiliaries can, however, be used as hedges in conjunction with both constructions (e.g. *he could possibly be right* vs. *it might become clear that that is the case*); since modal auxiliaries are not uniquely used in or together with the introductory *it* pattern, they will not be discussed further here.

From the discussion above, it becomes clear that these knockout factors can at least partly explain the distribution of these constructions. It then follows from this finding that there are, in fact, ways in which the two constructions differ such that they cannot always be considered to be paraphrases of one another, as there is information that can be conveyed through the introductory *it* pattern that is more difficult to convey through disjuncts.

4.2.2 *The distribution of variants*

With the instances of the introductory *it* pattern which contain additional information removed (see Section 4.2.1 above), we can use a binomial generalized linear

model to test which of the remaining factors (lexical influence, level of expertise in academic writing and L1) significantly affect the distribution of the two constructions with greater precision. Only disjuncts and instances of the introductory *it* pattern with *that*-clause complementation that did not include any of the features discussed above (such as *it is interesting that* and *interestingly*) were thus included in the model. There were 720 such tokens in the data. Since the frequencies in the present data strongly suggest that the present tense is the default tense marking for instances of the introductory *it* pattern, the past tense tokens were considered marked (and thus as representing the choice between present and past tense that is only possible for the introductory *it* pattern); the past tense tokens were therefore excluded before fitting the model.

When a model was fitted onto the data it became clear that L1 is not a significant predictor;⁹ lexis (base form) and expertise in academic writing were, by contrast, both significant, as was one of the interactions between the two (that of *possib**). The output of the binomial generalized linear model is shown as a model where the non-significant main predictor (L1) has been excluded in Table 2 below. This type of model enables tests for statistical significance of categorical variables and is thus of great use here, as all the variables are categorical rather than numeric.

What these results show is that compared to the experts, the learners make significantly less frequent use of disjuncts overall in relation to the introductory *it* pattern, as can be seen from the negative *z* value. Conversely, the experts make less frequent use of the introductory *it* pattern in relation to disjuncts. One potential reason for the experts' less frequent use of the introductory *it* pattern could be that they have been affected by style guides taking a prescriptive stance against this construction, arguing that the pattern makes the writing more wordy (cf. e.g. *The Bedford Handbook*, Hacker & Sommers 2014). However, the corpora used in this study does not allow for further investigation of whether the expert writers' use of the pattern has been affected by style guides; this will therefore have to be left for future studies to explore.

Furthermore, there is statistically significant inter-lexical variation. Using *clear** as the baseline, *possib** is significantly less likely to be used in the disjunct construction, whereas *importan** and *interest** are significantly more likely to be realized as disjuncts. Furthermore, the interaction between the base form *possib** and level

9. Differences across L1s were tested rather than L1 transfer, as the latter is a complex and elusive variable that is not easy to test statistically. The fact that there is no statistically significant difference between the L1 Swedish data and the L1 French data means that there is not enough evidence that the two learner groups behave differently from each other, which does not preclude L1 transfer. See Herriman (2013) for a more detailed discussion of the use of the introductory *it* pattern in Swedish.

Table 2. Output from the binomial generalized linear model comparing disjuncts and instances of the introductory *it* pattern that do not include any additional information

Coefficients	Estimate	Std. Error	<i>z</i> value	<i>p</i> (> <i>z</i>)
(Intercept)	0.99	0.17	5.67	1.47e ⁻⁰⁸ ***
LEVEL_Learner	-1.10	0.26	-4.20	2.63e ⁻⁰⁵ ***
BASE_FORM_IMPO	1.92	0.62	3.11	0.0019 **
BASE_FORM_INTE	2.38	0.74	3.22	0.0013 **
BASE_FORM_POSS	-0.80	0.29	-2.72	0.0065 **
BASE_FORM_SURP	-0.70	0.42	-1.67	0.0948 .
LEVEL_Learner:BASE_FORM_IMPO	-0.71	0.77	-0.92	0.3573
LEVEL_Learner:BASE_FORM_INTE	0.46	0.92	0.50	0.6167
LEVEL_Learner:BASE_FORM_POSS	1.55	0.41	3.77	0.0002 ***
LEVEL_Learner:BASE_FORM_SURP	1.12	0.58	1.93	0.0539 .

Nagelkerke's R^2 : 0.205

C: 0.776

G: 111.88; d.f. 9; $p < 0.001$ ¹⁰

of expertise in academic writing (as compared to *clear**) is statistically significant, meaning that their joint effect on the dependent variable (i.e. type of construction) “is not predictable from their individual effects on the same dependent variable” (Gries 2013: 249). It can thus be concluded that there is clear inter-lexical variability and that the learners are not using the two constructions in a fully expert-like manner.

An overview of the interactions is shown in Figure 6. The five base forms are displayed on the x-axis and the probability for the disjunct construction, given one of the base forms, is shown on the y-axis; the closer to 1, the more likely this base form is to be realized as a disjunct. The Effects package (Fox & Hong 2009), which is used here to display the results, not only provides an overview, but also provides the confidence intervals, which are marked by the whiskers. The confidence intervals show how likely it is that the true value lies within the interval produced by the model. The dashed line marks the 0.5 level.

10. Nagelkerke's R^2 , which is a measurement of variability, ranges from 0 to 1 (higher values indicate a better model fit) (Gries 2013: 265); the *C*-value, which shows the classification quality of the model, ranges from 0.5 to 1 (values over 0.8 are considered good) (Gries 2013: 304); *G* gives the results of a model likelihood ratio test, which shows that the model is statistically significant at the 0.001 level.

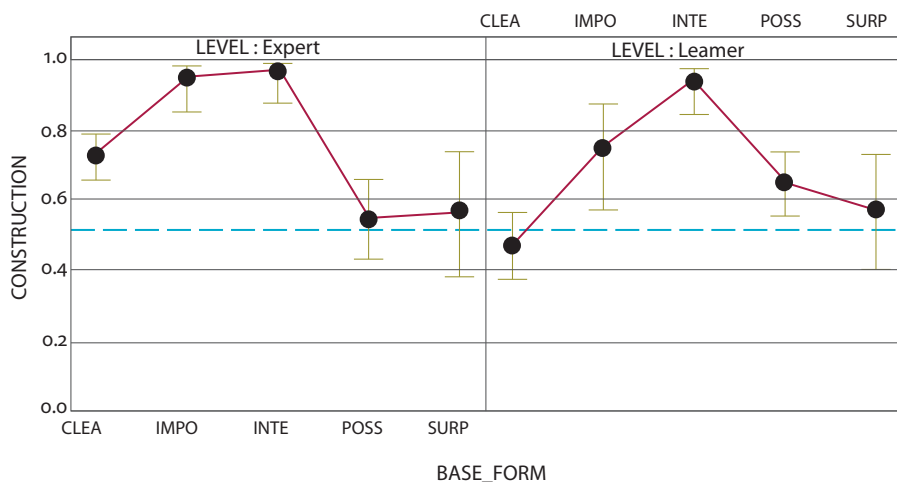


Figure 6. Effect plot showing the interaction between base form and level of expertise in academic writing, with the instances of the introductory *it* pattern containing added information removed

As can be noted, the confidence intervals are relatively large. Yet, three out of five of the base forms in the expert data (*clear**, *important** and *interest**) and three of the base forms in the learner data (*important**, *interest** and *possib**) are found to be significantly more likely to be realized as disjuncts than through the introductory *it* pattern, as these are found to be (fully) above the 0.5 line separating the likelihood of a disjunct from the likelihood of an instance of the introductory *it* pattern. Had the constructions been equally likely to occur (i.e. had the likelihood been 0.5 for all five of the base forms), the factors that are now removed would have accounted for all that makes the two constructions differ. However, as it is, there are still clear (and statistically significant) differences between the base forms, as well as across level of expertise in academic writing, which suggests that these are important factors that, along with the other factors discussed in Section 4.2.1, affect the distribution of the constructions.

5. Conclusion

The present study has investigated morphologically related stance markers that can be realized through different constructions. When the realizations of stance marking were examined in the first part of the study, it became clear that all three factors investigated (lexical influence, expertise in academic writing and L1 transfer) were important. First, there was clear inter-lexical variability between the different base

forms. From this perspective, future studies should take care not to compare constructions or syntactic categories without taking lexis into account, as witnessed by the fact that each of the five high-frequency base forms had a different behavioral pattern. Second, the learners tended to overuse almost all stance markers in relation to the experts, which is in line with previous research. In particular, the learners tended to use the base forms *interest** and *probab** to a much larger extent than the experts did. Third, not only did the learners overuse *interest** in general, they also significantly underused the adverbial form, *interestingly*, compared to the experts. Since *interestingly* does not have a direct equivalent in Swedish or French, the less frequent use of *interestingly* on the part of the learners could be explained by negative L1 transfer. The competing construction, the stance complement clause construction (e.g. *it is interesting that*), does, by contrast, exist in both languages, which might suggest that it is a safer choice for the learners. Other differences that might be due to L1 transfer include the L1 French students' underuse of *presumably*, as well as the L1 Swedish students' slight tendency to place *possibly* sentence-initially.

In the second part, in which the two most similar constructions, disjuncts and the introductory *it* pattern were compared, further differences were found that could help explain their distribution. For example, the introductory *it* pattern gives writers the option of adding information through certain adverbs and pre-clausal hedges. Along with these differences noted between the two constructions, two of the three factors previously investigated, lexis and level of expertise in academic writing, proved to be important predictors. First, there were statistically significant differences between the different base forms; for example, compared to *clear**, *interest** was more likely to be realized as a disjunct than as an instance of the introductory *it* pattern. Second, compared to the experts, the learners were found to have a greater tendency to use the introductory *it* pattern rather than disjuncts.

All in all, there seem to be principled explanations for why one grammatical realization of stance is used instead of another. However, as has been shown, the predictors form an intricate network, and more research is needed to explore further what affects the use of different stance markers and stance-marking constructions. Such studies could, for example, use POS-tagged (and/or parsed) corpora to carry out more large-scale comparisons across disciplines and/or genres. The present study can be seen as the first step towards a more complete description of what influences the distribution of constructions or realizations, as it allows for a macro perspective on some of the different kinds of stance marking available to writers. The study has moreover tried to detect problem areas for learners by comparing learner and expert writing, which can hopefully be useful for EAP teaching and theory. With regard to teaching, the results of the study could be utilized to help students learn to make more varied use of stance markers. With regard to theory,

in showing that some of the differences between the learners and experts seem to be L1-specific, the study highlights the importance of including more than one L1 in such investigations.

Acknowledgements

I would like to extend my warmest thanks to Merja Kytö and Erik Smitterberg for their invaluable comments during the course of this project. I am also grateful to the participants at higher seminars at the Department of English at Uppsala University and at the Centre for English Corpus Linguistics at Université catholique de Louvain for their helpful feedback on the project. Finally, I would like to thank the editorial team and the anonymous reviewers who made many valuable suggestions for improving the article.

References

- Anthony, L. (2014). *AntConc (Version 3.4.3)* [Computer Software]. Tokyo: Waseda University. Retrieved from <http://www.laurenceanthony.net/> (last accessed February 2017).
- Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116. doi:10.1016/j.jeap.2006.05.001
- Biber, D. (2006b). *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.23
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Charles, M. (2003). ‘This mystery...’: A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2(4), 313–326. doi:10.1016/S1475-1585(03)00048-1
- Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 56–73). Oxford: Oxford University Press.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 520 Million Words, 1990-present*. Available online at <http://corpus.byu.edu/coca/> (last accessed February 2017).
- Fox, J., & Hong, J. (2009). Effect displays in R for Multinomial and Proportional-odds logit models: Extensions to the Effects package. *Journal of Statistical Software*, 32(1), 1–24. doi:10.18637/jss.v032.i01
- Gries, S. T. (2013). *Statistics for Linguistics with R: A Practical Introduction*. Berlin: Mouton de Gruyter. doi:10.1515/9783110307474
- Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, 4(3), 257–277. doi:10.1016/j.jeap.2005.03.002
- Hacker, D., & Sommers, N. I. (2014). *The Bedford Handbook* (9th ed.). Boston, MA: Bedford/St. Martins.
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd ed.). London: Edward 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* (pp. 121–139). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.33.12has
- Hatzitheodorou, A.-M., & Mattheoudakis, M. (2009). “It is more than true that television reproduces life”: The effect of Greek rhetorical conventions on Greek learners’ academic writing in English. In T. Tsangalidis (Ed.), *Selected Papers from the 18th International Symposium on Theoretical and Applied Linguistics* (pp. 167–176). Thessaloniki: Monochromia.
- Herriman, J. (2013). The extraposition of clausal subjects in English and Swedish. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus-based Contrastive Linguistics: Studies in Honor of Stig Johansson* (pp. 233–260). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.54.13her
- Herriman, J., & Boström Aronsson (2009). Themes in Swedish advanced learners’ writing in English. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 101–120). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.33.11her
- Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/scl.4
- Hunston, S., & Sinclair J. (2000). A local grammar of evaluation. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial Stance and the Construction of Discourse* (pp. 74–101). Oxford: Oxford University Press.
- Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173–192. doi:10.1177/1461445605050365
- Koehn, P. (2005). Europarl: A parallel corpus for statistical machine translation. In *Conference Proceedings: The Tenth Machine Translation Summit (MT Summit X)* (pp. 9–86). Phuket, Thailand.
- Larsson, T. (2014, January). *Introducing the Advanced Learner English Corpus (ALEC): A new learner corpus*. Poster presented at the 2014 LOT Winter School, Amsterdam, The Netherlands.
- Larsson, T. (2016). The introductory *it* pattern: Variability explored in learner and expert writing. *Journal of English for Academic Purposes*, 22(2016), 64–79.
- Larsson, T. (forthcoming). A syntactic analysis of the introductory *it* pattern in non-native-speaker and native-speaker student writing. In M. Mahlberg & V. Wiegand (Eds.), *Corpus Linguistics, Context and Culture*. Berlin: De Gruyter Mouton.
- Larsson, T. (under review). A functional classification of the introductory *it* pattern: Investigating academic writing by non-native-speaker and native-speaker students.
- Lindgrén, S.-A. (2015). *The BATMAT Corpus Version 1.1*. English Language and Literature, Åbo Akademi University, Åbo.
- LOCRA (Louvain Corpus of Research Articles) [Corpus]. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université catholique de Louvain. Available online at <https://www.uclouvain.be/en-cecl-locra.html> (last accessed February 2017).
- Martin, J. R., & White, P. R. R. (2005). *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
- Neff, J., Ballesteros, F., Dafouz, E., Martínez, F., & Rica, J.-P. (2003). Formulating writer stance: A contrastive study of EFL learner corpora. In D. Archer, P. Rayson, A. Wilson & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference. UCREL Technical Papers 16* (pp. 562–571). Lancaster University: UCREL.

- Paquot, M., Hasselgård, H., & Oksefjell Ebeling, S. (2013). Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In S. Granger, G. Gilquin & F. Meunier (Eds.), *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead* (pp. 377–388). Louvain-la-Neuve: Presses universitaires de Louvain.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- R Core Team (2015). R: A language and environment for statistical computing [Computer software]. Vienna, R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/> (last accessed November 2016).
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Appendix 1. Raw frequencies of the base forms and the three grammatical realizations of stance included in the study in the three corpora

	Stance complement clause construction	Stance adverbs	Stance noun + prep. phrase	Total
Experts	301	285	184	770
<i>importan*</i>	92	55	87	234
<i>interest*</i>	19	58	10	87
<i>possib*</i>	177	45	72	294
<i>presum*</i>	3	61	3	67
<i>probab*</i>	10	66	12	88
Learners L1: Swedish	398	414	159	971
<i>importan*</i>	119	22	102	243
<i>interest*</i>	72	57	15	144
<i>possib*</i>	201	70	37	308
<i>presum*</i>	3	50	1	54
<i>probab*</i>	3	215	4	222
Learners L1: French	36	37	7	80
<i>importan*</i>	10	2	4	16
<i>interest*</i>	20	4	1	25
<i>possib*</i>	6	1	1	8
<i>presum*</i>	0	1	0	1
<i>probab*</i>	0	29	1	30
Total	735	736	350	1,821

Appendix 2. Summary of the coefficients and standard errors for a multinomial log-linear model fitted onto the results of the frequencies for each base form across the corpora investigated

Coefficients	Learners		
	(Intercept)	L1: French	L1: Swedish
<i>interest*</i>	-0.989	1.436	0.454
<i>possib*</i>	0.232	-0.925	0.038
<i>presum*</i>	-1.251	-1.522	-0.266
<i>probab*</i>	-0.978	1.607	0.875

Std. Errors	Learners		
	(Intercept)	L1: French	L1: Swedish
<i>interest*</i>	0.126	0.344	0.164
<i>possib*</i>	0.088	0.442	0.122
<i>presum*</i>	0.139	1.040	0.204
<i>probab*</i>	0.125	0.334	0.156

Author's address

Tove Larsson
 Department of English
 Uppsala University
 Box 527
 SE-75120 Uppsala
 Sweden
 tove.larsson@engelska.uu.se