

L1 novice writing as a missing piece in the Learner Corpus Research puzzle: The case of hedging

Pauline Jadoulle

Université catholique de Louvain

This paper investigates the possible benefits of integrating an L1 novice corpus into the study of learner writing, through a study on hedging adverbials and complement clauses. A more traditional tripartite comparison is first drawn between French EFL learner writing, L1 English novice writing, and English expert writing. Then, a comparison is made between L1 French novice writing and French EFL learner writing. The study uses data from the following corpora: the *Varieties of English for Specific Purposes dAtabase* (VESPA; Paquot et al., 2022) (French EFL learner writing), the *British Academic Written English* (BAWE; Heuboeck et al., 2008) corpus and the *Michigan Corpus of Upper-level Student Papers* (MICUSP; Ädel & Römer, 2012) (both L1 English novice writing), the *Kulturell Identitet i Akademisk Prosa: nasjonal versus disiplinavhengig* (KIAP-EN; Fløttum et al., 2006) corpus (English expert writing), and the *The French Academic Writing* (FAR) corpus (L1 French novice writing). The results speak to the importance of including a L1 novice component to the current methodologies employed in Learner Corpus Research (LCR), since such inclusion helps to provide a better and more nuanced interpretation of the findings obtained by means of a more traditional tripartite approach.

Keywords: L1 novice writing; learner writing; Learner Corpus Research; hedging

1. Introduction

In the past decades, the field of Learner Corpus Research (LCR) has shed light on typical features of learner academic writing, especially English as a Foreign Language (EFL) learner writing. For example, EFL learners have extensively been documented to show an involved style (e.g. Crosthwaite et al., 2017; Jiang, 2015; Wang & Jiang, 2018), an assertive tone (e.g. Hyland & Milton, 1997; Hyland & Tse, 2005; Paquot, 2010), or a limited lexical repertoire (e.g. Ädel & Erman, 2012; Larsson et al., 2022; Salazar, 2014; Schweinberger, 2020; Staples & Reppen, 2016). Those findings are largely based on two types of studies. LCR studies on English academic writing usually conduct pairwise comparisons between English native novice writers and EFL learners (e.g. Chen, 2012; Ebeling & Hasselgård, 2015; Hyland & Milton, 1997; Jiang, 2015; Lee & Deakin, 2016; Neff et al., 2003; Schweinberger, 2020; Shin, 2019; Staples & Reppen, 2016), or between English expert writers and EFL learners (Hyland & Tse, 2005; Larsson, 2016; Paquot, 2010; Salazar, 2014; Su et al., 2022; Wang & Jiang, 2018; Wei & Lei, 2011). Both these approaches have been criticized: the former because native novice writing does not necessarily constitute a standard for learners to reach (Leech, 1998: xix), and the latter because expert writing is said to be “both unfair and descriptively inadequate” as a reference model (Lorenz, 1999: 14). Against this backdrop, Ädel (2006: 207) argued that “the ideal case [...] would be to use both native-speaker [texts] and professional texts as standards for comparison” so as to reach a double objective, i.e. to meet the principle of ‘sameness’ and to have a professional reference model. In line with this suggestion, another approach emerged, which consists in carrying out tripartite comparisons between English experts, English native novices, and EFL learners (e.g. Chen & Baker, 2010;

Larsson, 2019; Larsson et al., 2022; Taqvist, 2018). Although one cannot deny the considerable contributions of these three approaches to the field of LCR, expanding these ‘traditional’ LCR methodologies could help us gain a better understanding of the complex interplay of factors influencing EFL learner writing.

In fact, what seems to be missing from these three approaches – and could help LCR linguists get better insights into this complex interplay of factors – is the inclusion of a corpus of novice writing in the L1 of EFL learners. This is precisely what I propose to include in the present study. Theoretically, several language transfer models including an L1 novice component have been put forward. Jarvis (2000) includes the learners’ L1 in his framework for the study of L1 influence, and mentions that robust correspondence between the latter and learner writing “may be the strongest type of evidence for L1 influence” (Jarvis, 2000: 258). Gilquin (2008) also includes an L1 native novice variety in one of the comparisons of her Detection-Explanation-Evaluation (DEE) model designed to study transfer (namely NL vs. IL). Such a comparison remains to be implemented, however, as almost no LCR study has looked at L1 novice writing in combination with EFL learner writing. One notable exception is Çandarlı et al. (2015), who analyzed authorial presence markers in Turkish EFL learner writing, Turkish L1 novice writing, and English L1 novice writing¹. Using a corpus of L1 novice writing allowed the researchers to better understand possible influences behind EFL learners’ use of authorial presence markers: the authors outline features impacted by cultural tendencies (e.g. first-person pronoun singular, whose frequency is similar across L1

¹ See also Ströbel et al. (2020) for a study on the relationship between L1 and L2 language complexity in novice writing by the same individuals, Dirdal (2022) for a small-scale comparison of relative clauses in Norwegian EFL schoolchildren, L1 Norwegian writing by the same children, and L1 English writing by British schoolchildren, and Rørvik (2021) for a comparison between Norwegian and English writing produced by presumed L1 Norwegian students (‘presumed’ as the L1 of the writers could be not identified with certainty in the study).

Turkish and L2 English writing by Turkish students) and features impacted by English language conventions (e.g. boosters, whose frequency is similar across L2 English writing and L1 English writing, and much more frequent in L1 Turkish texts).

Whether adding a native novice corpus in the learners' L1 background to the LCR methodological toolbox actually enhances our understanding of L2 English writing has not been systematically examined however, and we need empirical studies to highlight the importance of collecting and using this type of data. The present study aims to fill this gap by answering the following research question: "To what extent does adding a corpus of novice writing in the L1 of the learners complement the LCR methodological apparatus and contribute to a better understanding of the complex interplay of factors that influence learner language?". This question will be answered by means of a corpus-based study of hedging in French EFL learner academic writing.

Research has shown that EFL learners from different L1 backgrounds tend to underuse hedges compared to both English experts and L1 English novice writers, resulting in an overly forceful and assertive tone (Hinkel, 2005; Hyland & Milton, 1997; Hyland & Tse, 2005; Larsson, 2017; Paquot, 2010). Chen (2010), for instance, found a higher frequency of hedges like *may* and *possible* in expert writing as compared with learner writing. Similar findings have been reported by Larsson (2019), who compares L2 English novice writers, L1 English novice writers, and expert English writers. She observes that "while neither the NS students' nor the learners' use of hedges is identical to that of the academic experts, the NS students' use is slightly closer overall" (Larsson, 2019: 259). In addition, EFL learners appear to tap into a more limited lexical repertoire in comparison with their native and expert counterparts, therefore exhibiting difficulties in using a wide and nuanced range of hedging expressions (Chen, 2012; Hinkel, 2005;

Hyland & Milton, 1997; Hyland & Sancho Guinda, 2012; Jiang, 2015; Wang & Jiang, 2018). Jiang (2015), for example, notes this lack of variety in EFL learners' use of hedging nouns. The author qualifies the learners' use of 'teddy bears' like *opinion*, *idea*, *view* and *thought* as "pragmatically vague and void of illocutionary force" (Jiang, 2015: 96), in contrast with L1 students who use nouns like *argument* and *claim*, which directly support their argumentation. Those features of learner writing (i.e. an underuse of hedges, a more limited repertoire of hedging devices) could be influenced by a variety of factors, such as L1 transfer. Albeit often assumed to influence learners' use of hedges, L1 transfer is often mentioned in passing and as a hypothetical reason behind a particular tendency observed, in the discussion or conclusion section (e.g. Chen, 2010; Jiang, 2015). It is often based on researchers' intuition or interpretation. The present study sets out to investigate whether the influence of the learners' L1 is validated by empirical data.

2. Methodology

This paper follows studies like Larsson (2019) and Römer (2009) and distinguishes between 'novice' (or 'apprentice') and 'expert' writing, with novice writing being defined as "unpublished pieces of writing that have been written in educational or training settings, (often) for purposes of assessment" (Scott & Tribble, 2006: 133) and 'expert' writing as pieces of writing published in peer-reviewed journals. In line with our objective to investigate the possible advantages of integrating an L1 component to LCR methods, a two-step methodology is adopted. First, a tripartite comparison is drawn between a corpus of English expert writing, a corpus of L1 English novice writing, and a corpus of

French EFL learner (or non-native novice) writing. This step replicates commonly used methods in LCR, and thus seeks to gain insights into French EFL learner language in a more ‘traditional’ way. Following this, a comparison is made between the same corpus of French EFL learner writing and a corpus of L1 French novice writing. This step will help us understand the extent to which the addition of a L1 novice corpus contributes to a better understanding of the complex interplay of factors influencing learner writing. The aforementioned objective is achieved by means of a study of hedges, involving its own methodological implications (e.g. operationalization, identification).

2.1 Corpus data

This section briefly describes the four corpora analyzed, and discusses their comparability.

2.1.1 French EFL learner writing

The VESPA-LING-FR (henceforth VESPA) is a sample from the *Varieties of English for Specific Purposes dAtabase* (Paquot et al., 2022). The corpus contains English term papers written by French-speaking students in the framework of linguistics courses. The proficiency level of those students ranges from upper-intermediate to advanced. The corpus comprises 98 texts totaling 371,422 words. Forty-eight texts were written by BA students and 50 texts by MA students.

2.1.2 L1 English novice writing

To ensure comparability, the corpus of L1 novice writing used in this study also contains texts from the discipline of linguistics. Following Larsson (2016), and to include a sufficient number of texts, data from two corpora were combined, namely the *British Academic Written English* (BAWE; Heuboeck et al., 2008) and the *Michigan Corpus of Upper-level Student Papers* (MICUSP; Ädel & Römer, 2012). The corpus therefore comprises both British and American texts. It will be referred to as the ‘BAWE & MICUSP corpus’.

The texts selected from the BAWE are linguistics papers written by (L1) English-speaking students enrolled at British universities at four levels of study, from undergraduate to MA level. They can be qualified as samples of successful writing, as they meet the departmental requirements for the corresponding study level (Sian & Nesi, 2009). These texts are meant to be evaluated and are written so that students “demonstrat[e] academic knowledge and skills as opposed to present research findings” (Nesi, 2008: 239). For comparability purposes, and following Larsson (2016), the texts classified in the ‘research report’ and ‘essay’² categories are included.

The texts selected from the MICUSP are A-graded student linguistics papers written by native speakers of English studying at the University of Michigan. The texts were written by senior undergraduate and second-year graduate students (Ädel & Römer, 2012; Römer & O’Donnell, 2011). Following Larsson (2016), the texts categorized as ‘research papers’, ‘argumentative essays’³ and ‘proposals’ are included.

² In the BAWE manual, ‘essay’ is not used in the sense of ‘argumentative essay’ (see Heuboeck et al., 2008).

³ Argumentative essays refer to discipline-specific, thesis-driven papers requiring reference to external sources (O’Donnell & Römer, 2012). Examples of argumentative essays from the MICUSP are “National Identity and Language Education Policy” or “A Discussion of the VP-External and VP-Internal Subjects Hypotheses”.

The BAWE & MICUSP corpus contains 73 texts and amounts to 175,677 words (c. 53 texts with 115,000 words from the BAWE, and 20 texts with c. 60,000 words from the MICUSP).

2.1.3 English expert writing

The KIAP-EN-LING (henceforth KIAP-EN) is a sample from the *Kulturell Identitet i Akademisk Prosa: nasjonal versus disiplinavhengig* (Fløttum et al., 2006). It contains 50 research articles written in English in the discipline of linguistics, and amounts to 530,809 words. The texts are mainly expository in nature and “primarily report and discuss empirical (including experimental) or theoretical research that has been done by the authors themselves” (Fløttum et al., 2006: 8). The authors of these texts are expert academic writers⁴, that is, professional academics whose papers were published in peer-reviewed journals.

2.1.4 L1 French novice writing

The *French Academic Writing* (FAR) corpus is a corpus of native novice French academic writing. It is comprised of academic papers written by French-speaking students in the framework of their linguistics courses at Université catholique de Louvain (UCLouvain, Belgium). It contains 70 texts amounting to 370,353 words. Nineteen papers were written by third-year BA students and 51 papers by MA students. Examples of paper topics are presented in Table 1.

⁴ In their description of the KIAP corpus, Fløttum et al. (2006: 9) state that “[a]s far as possible, texts have been chosen where the author (or at least one of the authors) appears to be a native speaker of the language in which the article is written (on the basis of his or her name and other information available)”. The authors are therefore described as “presumed native speakers” (Fløttum et al. 2006: 7).

Table 1. Paper topics from the FAR

FILE	TOPIC
FAR0006	Enquête sociolinguistique. La perception du genre linguistique chez les enfants de sixième primaire. [EN] Sociolinguistic survey. The perception of linguistic gender among sixth grade children.
FAR0028	Le rôle de la langue maternelle dans l'apprentissage d'une langue seconde : Les erreurs « typiques » commises par les apprenants hispanophones dans leur apprentissage du français langue étrangère. [EN] The role of the mother tongue in second language learning: "Typical" errors made by Spanish-speaking learners in French as a foreign language.
FAR0072	Etude de corpus publicitaires [EN] A study of advertising corpora
FAR0080	La vulgarisation scientifique dans les documentaires [EN] Popularization of science in documentaries
FAR0097	Etude de l'influence de la situation de communication sur le phénomène de disfluence en français parlé. Analyse de la vitesse de parole dans la production de disfluences. [EN] The influence of the situation of communication on disfluency in spoken French. Analysis of the speed of speech in the production of disfluencies.

2.1.5 Comparability of the corpora

The texts included in the different corpora present many commonalities, as they mainly consist in empirical studies carried out in the field of linguistics. The term papers from the VESPA were collected in the framework of linguistics courses like Lexicology or Corpus Linguistics, and analyze specific words or constructions in certain registers, language varieties, etc. For instance, the text UCL0035-LING-01 investigates the use of the Saxon genitive in music articles. The papers from the KIAP-EN are quite similar to the VESPA term papers, as they share the same primary goal and focus. They typically involve in-depth analyses of specific linguistic phenomena, constructions, or items, which are studied from specific semantic, syntactic, etc. perspectives (e.g. “light verbs and polysemy”, “determiners in genitive constructions”). The FAR was built to be as comparable as possible to the VESPA. Its texts report on linguistics research based on empirical or theoretical data (see examples of titles above). ‘Research reports’ and ‘research papers’ from the BAWE & MICUSP are also similar to ‘term papers’ from the VESPA, as they seek to “demonstrate ability to undertake a complete piece of research including research design, and an appreciation of its significance in the field” (Heuboeck et al., 2008: 47). Text 6020b is a study of language before and after age 5, for instance, and text LIN.G0.10.2 deals with “The Distribution of Anaphoric ‘So’ in Micase”.

However, a number of register and text type differences should be acknowledged. This limitation mainly concerns the BAWE & MICUSP corpus. In their ‘essays’, BAWE students are asked “to develop the ability to construct a coherent argument and develop critical thinking skills” (Heuboeck et al., 2008: 47). Such texts present differences compared to those of the other corpora analyzed, in that they deal with broader linguistic concepts and contexts, like gender variation in speech and dialect acquisition, and often

require the writers to discuss specific questions or prompts (e.g. What is the effect of the social variables of class and gender on the way we speak?). Those texts therefore have a more persuasive dimension than the text types of the KIAP-EN, VESPA, and FAR. ‘Argumentative essays’ and ‘proposals’ from MICUSP show a close resemblance to ‘essays’ from the BAWE, in that they also typically discuss broader linguistic concepts such as Language education policy or Language Teaching. The results of the present study will therefore be discussed in light of these caveats (see Section 4).

Secondly, due to different situational and communicative characteristics, published articles do not present the exact same register features as student papers written for assessment purposes. One of the main differences concerns the fact that students write to a ‘reader-evaluator’, which can impact the degree of caution they adopt in their texts. These differences should be taken into account when considering the results of the present analysis.

Finally, it should be acknowledged that the BAWE & MICUSP and the VESPA corpora contain a higher proportion of upper-level student writing compared to the FAR (about 50% of the texts in the BAWE & MICUSP and in the VESPA compared to about 30% in the FAR). In addition, the BAWE & MICUSP is made up of successful pieces of academic writing, while this is not necessarily the case for the two other corpora of novice writing. The discussion section will address the possible impact of this last difference on the results.

Most of these differences, albeit important to acknowledge, are indicators of ecological validity. The VESPA, BAWE & MICUSP, and FAR corpora contain texts that were naturally produced in the framework of linguistics courses at university. As Biber et al. (2020) argue, while such texts can present variation in terms of text type, register,

and content, they are also valuable as they represent authentic data from disciplinary courses. This is to be contrasted with texts produced in the framework of strictly controlled data collections, usually carried out in the framework of writing courses. In addition, BAWE & MICUSP texts come from Anglo-Saxon university settings while VESPA and FAR texts were collected in Belgium. Gathering perfectly comparable data from these two settings is not possible, as students in different cultural and university environments are often required to produce different types of texts, etc. Given the fact that ensuring perfect comparability is out of our reach, the challenge in the present paper will be to take these differences into account in interpreting the results.

2.2 Operationalization of hedging

In this paper, hedging is defined as the linguistic ways writers express caution, uncertainty, and distance towards the information presented, thereby interacting with an audience (Biber et al., 1999; Hyland, 2005). This function is conveyed through the use of markers like *maybe*, *hypothesis* or *probable* (Biber et al., 1999; Hyland, 2005). The present study focuses on metadiscursive markers, that is, markers that “overtly express an attitude or assessment relative to some proposition in the immediate discourse” (Gray & Biber, 2014: 221). In other words, I take into account markers that, in the same sentence, clearly qualify propositional content by means of two overt components, namely 1) a hedging marker, and 2) a proposition framed by this marker (Gray & Biber, 2014). This means that the markers *assume* or *hypothèse* [En. *hypothesis*] in sentences (1) and (2) are not included in the study. By contrast, the same markers used in (3) and (4) are included in the study.

- (1) *No direct explanation is given for this feature, but as we will see for other features, the influence of Chinese and Malay languages is to be **assumed**.* (VESPA, UCL1321-LING-01)
- (2) *Nous ne formulerons donc pas d'**hypothèse** sur base de ces données.* (FAR, 0012)
- ‘We will therefore not formulate any **hypothesis** based on this data’
- (3) *We may **assume** that this phenomenon would be linked to the audience targeted by the newspaper.* (VESPA, UCL1304-LING-01)
- (4) *On pourrait même émettre l'**hypothèse** que la dimension sociale de la féminisation serait un avantage en faveur de celle-ci.* (FAR, 0015)
- ‘One could even formulate the **hypothesis** that the social dimension of feminization would be an advantage in favor of it’

On this basis, three main grammatical devices can be said to convey hedging, i.e. adverbials (e.g. *maybe*), complement clause constructions controlled by adjectives, nouns, and verbs (e.g. *likely/assumption/claim* + complement clause), and modal verbs (e.g. *may*) (Biber et al., 2018)⁵. This paper focuses on adverbials and complement clauses controlled by adjectives, nouns, and verbs⁶.

⁵ ‘Stance noun + prepositional phrase’ constructions are also used to express hedging (e.g. Biber et al., 2018).

⁶ For time and space reasons, modal verbs were excluded from the present study. Modal verbs constitute a separate area of study, with its own literature and theories. They require a great deal of disambiguation, with complex guidelines to be established downstream (e.g. Vold, 2008).

2.3 Hedge identification

Studies on stance and hedging in academic writing typically work on the basis of lists of lexico-grammatical markers (e.g. Biber, 2006b; Hyland, 2005; Yang, 2013). Hyland's (2005) list, for example, comprises 101 hedges, including *claim*, *probably* and *plausible*. Most of these lists contain English markers however, and, to my knowledge, no large-scale empirical study to date has designed a list of French hedges. Working with pre-existing lists would therefore not have allowed me to coherently identify and extract markers in both English and French. Instead, a methodology was developed to identify and extract hedges on the basis of their grammatical category. The aim was not to create an exhaustive collection of markers that might possibly appear in the corpora under study, but rather to directly search for the markers that are in fact used in these corpora. More concretely, adverbials were identified by extracting adverbs and n-grams from the corpora, and then selecting items that are considered as hedges. Complement clauses were accessed through an initial selection of controlling adjectives, nouns and verbs that occur in constructions like 'ADJ + *that*' or 'VERB + *que*'. This means that structures like *it is possible that* and *on suppose que* [En. *it is assumed that*] were found through the extraction of the controlling words *possible* and *supposer* [En. *assume*]. This section describes the identification process in more detail.

2.3.1 Adverbials

The category of adverbials includes single-word adverbs (e.g. *maybe*) and multi-word adverbial expressions (e.g. *sans doute* [En. *arguably*]). Adverbs were accessed by extracting the tags [tag="RB.?"] and [tag=".*R"] from the English and French corpora

respectively in *Sketch Engine* (Kilgariff et al., 2014)⁷. The items extracted were subjected to two criteria, which they had to meet to be retained. For an item to be included, it needed to 1) meet the semantic definition of hedging provided above, and 2) be used metadiscursively. This was verified by looking at the items in context. Multi-word adverbial expressions were identified separately by extracting 2- and 3-grams (minimum frequency = 3) from the English and French corpora. Five of the resulting 79,287 English n-grams and four of the 51,855 French n-grams were retained. The adverbials yielded by this method are presented on pages 17-18 (with the rest of the markers).

2.3.2 Complement clauses

Hedging complement clauses were found via their controlling adjectives (e.g. *apparent*), nouns (e.g. *possibilité* [En. *possibility*]), and verbs (e.g. *claim*). The main steps undertaken to find these controlling words are as follows:

1. Selection of prominent adjectival, nominal and verbal hedging patterns (e.g. noun + *that*, verb + *que*)
2. Extraction of these patterns in the corpora under study
3. Selection of controlling words based on two criteria (e.g. *hypothesis*, *penser*)

The following paragraphs provide a brief description of the process for English and French respectively.

⁷ *Sketch Engine* automatically tags the corpora uploaded for POS. For the English data it uses a modified tagset of the default TreeTagger tagset, and for the French data it uses the French FreeLing part-of-speech tagset.

The first step thus consisted in selecting prominent hedging patterns. For English, these patterns were selected based on the literature (e.g. Biber, 2006b; Larsson, 2019). The following structures were kept: ADJ + *that|to*, NOUN + *that|to* and VERB + *that|to*. They constitute the established adjectival, nominal, and verbal complement clause constructions in the stance tradition (Gray & Biber, 2014). On this basis, the queries [tag="J.*"] [word="that|to"], [tag="N.*"] [word="that|to"] and [tag="V.*"] [word="that|to"] were applied to the three English corpora in *Sketch Engine*, thus revealing a variety of potential controlling words⁸. These searches yielded about 400 adjectives, more than 1,500 nouns and more than 600 verbs (types). I then went over those potential controlling words and applied the same two selection criteria as the ones adopted for hedging adverbials (see Section 2.3.1). This selection process resulted in a total of 13 adjectives, 31 nouns, and 33 verbs (see below).

As, to our knowledge, no large-scale corpus-based empirical studies report the possible patterns of hedging in French, it was not possible to identify common hedging patterns on the basis of the literature. These patterns were therefore identified directly from a large corpus of French academic writing, namely the French subpart of the *Louvain Corpus of Research Articles* (LOCRA-FR; Dupont 2019). The corpus contains 313 French research articles (c. 2 million words) written by expert writers, from top-ranked journals in five Humanities disciplines, namely anthropology, education, political science, psychology and sociology⁹. To find frequently used hedging expressions, 2- to

⁸ I did not specify any POS for *that* and *to* so as to avoid tagging mistakes, which could have made me miss a number of items.

⁹ Note that this French expert corpus is therefore not used as part of the main analysis. It is simply used to complete step 1 presented above. A large corpus was needed to complete this step, so as not to miss important hedging patterns (which means that I could not use a corpus like the KIAP-FR or the FAR for instance).

6-grams including common hedging adjectives, nouns and verbs¹⁰ (minimum frequency = 5) were extracted from the LOCRA-FR (see Table 2). The hedging patterns found by means of this method were the following: ADJ + *que|qu|de|d*, NOUN+*que|qu|de|d|selon* and VERB+*que|qu|de|d|INF*.

¹⁰ These were chosen on the basis of Kanté (2010, 2016, 2019), Vold (2008), the French equivalents offered in the *Louvain EAP Dictionary* (LEAD) (see functions ‘express possibility and certainty’, ‘express personal opinion’, and ‘quoting and reporting’) (Granger & Paquot, 2010), and the transdisciplinary lexical items gathered by the *Laboratoire de Linguistique et de Didactique des Langues Etrangères et Maternelles* (LIDILEM) from *Université Grenoble Alpes* (<https://scientext.hypotheses.org/lexique-transdisciplinaire-v1>).

Table 2. Extraction of n-grams with common hedges from the LOCRA-FR

HEDGE	GRAMMATICAL CATEGORY	EXAMPLES OF HEDGING N-GRAMS EXTRACTED	HEDGING PATTERNS FOUND
<i>POSSIBLE</i> [EN. POSSIBLE]	Adjective	Possible que, est possible que, il est possible que, possible qu', est possible que l'	Possible que Possible qu'
<i>PROBABLE</i> [EN. PROBABLE]	Adjective	Il est probable que, probable de, probable que la, probable qu'	Probable que Probable de Probable qu'
<i>HYPOTHÈSE</i> [EN. HYPOTHESIS]	Noun	L'hypothèse que, l'hypothèse selon laquelle, faire l'hypothèse que, l'hypothèse que les, hypothèse qu, on peut faire l'hypothèse que	Hypothèse que Hypothèse selon (laquelle) Hypothèse qu'
<i>IMPRESSION</i> [EN. IMPRESSION]	Noun	L'impression de, impression que, impression d'	Impression de Impression que Impression d'

<i>SUGGÉRER</i>	Verb	Suggérer que, suggère que, suggère que les, suggère qu, suggèrent que, suggèrent qu, résultats suggèrent que les, suggéré que, suggéré de	Suggérer que Suggérer qu' Suggérer de
[EN. SUGGEST, IMPLY]			
<i>ESTIMER</i>	Verb	Estimer que, estime que, estime qu, estiment que, estiment être, estiment qu	Estimer que Estimer qu' Estimer + INF
[EN. CONSIDER, BELIEVE]			

Patterns with adjectives, nouns and verbs were extracted from the FAR in *Sketch Engine*, with the following queries:

- [tag="A.*"] [word="de|d|que|qu"] (for adjectives)
- [tag="N.*"] [word="que|qu|de|d|selon"] (for nouns)
- [tag="V.*"] [word="que|qu|de|d"] and [tag="V.*" & lemma!="avoir|être|pouvoir|devoir|faire|sembler|paraître|faillir¹¹|falloir"] [tag="V.*"]¹² (for verbs)

The searches yielded about 500 adjectives, 1,600 nouns, and 600 verbs (types). After applying the same two selection criteria as for the English items, this method yielded a total of 5 adjectives, 9 nouns, and 21 verbs (see below).

An overview of the three steps to find controlling adjectives, nouns, and verbs is summarized in Figure 1:

¹¹ The program mistakenly considered *faillir* as the infinitive of *faut* and *faillie*.

¹² This corresponds to all verbs except the most frequent (semi-)modals – which were creating a lot of noise – followed by another verb.

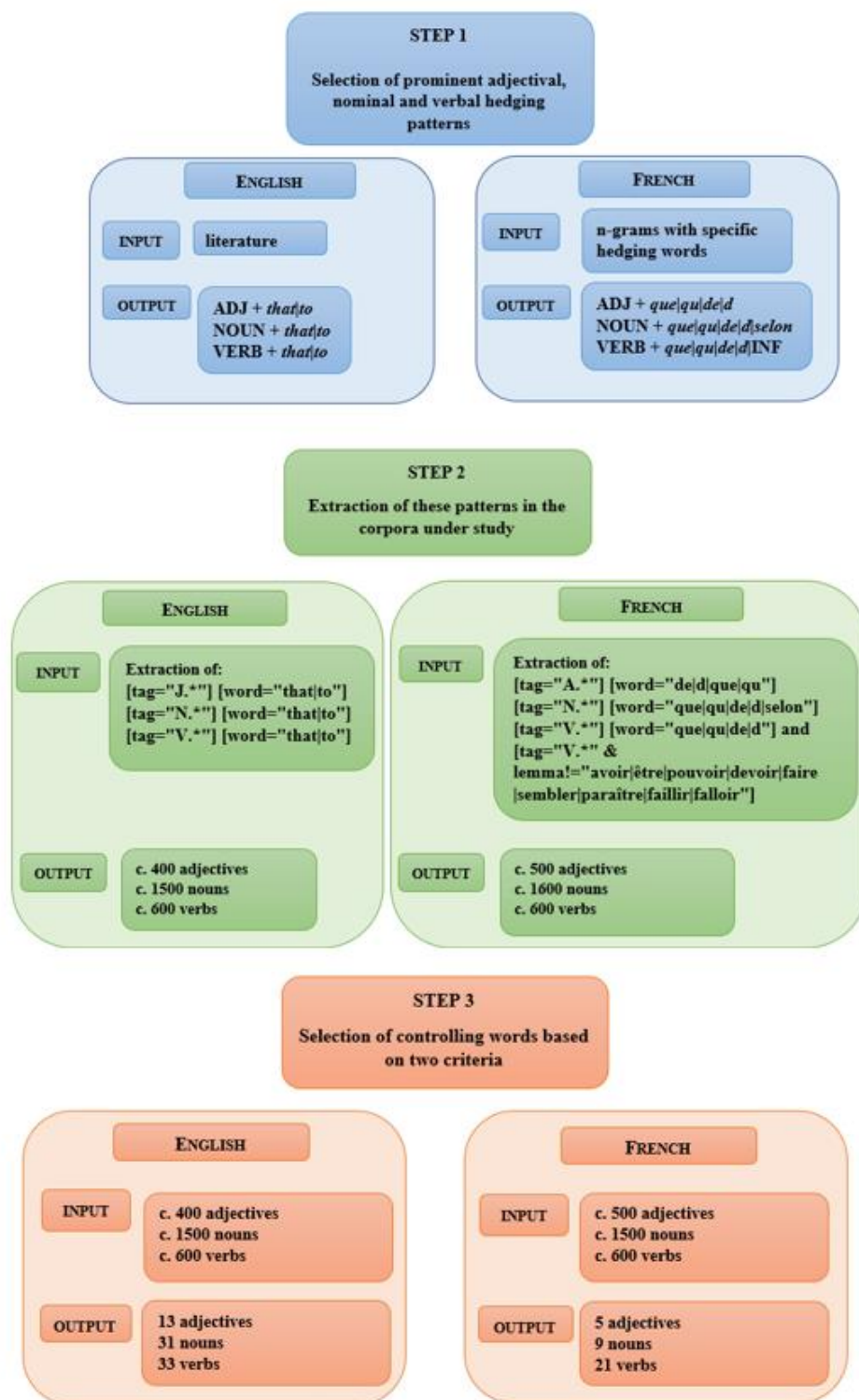


Figure 1. Main steps to find controlling hedging adjectives, nouns, and verbs

All the hedges selected can be found in Tables 3 (English) and 4 (French) below. Those hedges were extracted from the four main corpora under study (see Section 2.1). Note that the number of French hedges is much more limited than the number of English hedges. This is in accordance with previous research on epistemic stance in English and French academic writing, which showed the former to have a much stronger presence of hedging devices than the latter (Vold, 2008).

Table 3. English hedges under study

admittedly	perception
allege	perhaps
apparent	perspective
apparently	plausible
appear	plausibly
arguable	point
arguably	point of view
argue	posit
argument	position
assume	possibility
assumption	possible
belief	possibly
believe	postulate
chance	potential
claim (n.)	potentially
claim (v.)	presumably

conceivable	presume
conceivably	presumption
conjecture (n.)	probability
conjecture (v.)	probable
consider	probably
debatable	propose
deem	purport
doubt (n.)	purportedly
doubt (v.)	see
doubtful	seem
eventually	seemingly
feel	speculate
feeling	standpoint
find	suggest
guess	suggestion
hypothesis	suggestive
hypothesis/ze	suppose
imagine	supposedly
imply	supposition
impression	surmise
in theory	theoretically
infer	theory
interpret	thesis
interpretation	think

intuition	thought
likelihood	to my knowledge
likely (adv.)	to our knowledge
likely (adj.)	unclear
maybe	unlikely
not necessarily	unsure
not obligatorily	venture
opinion	view (n.)
ostensibly	view (v.)
perceive	viewpoint

Table 4. French hedges under study

apparaître	possiblement
apparemment	postulat
argument	postuler
avis	potentiellement
certes	présumer
chance	probable
conjecturer	probablement
considérer	proposer
croire	ressentir
décréter	sans doute
enclin	sembler
estimer	sentir
éventuellement	sous-entendre
hésiter	soutenir
hypothèse	suggérer
imaginer	supposément
impression	supposer
intuition	supposition
paraître	susceptible
pas forcément	théoriquement
pas nécessairement	trouver
penser	visiblement
peut-être	voir

possibilité

vraisemblable

possible

2.4 Data disambiguation

Once the identified markers were extracted, the occurrences needed to be disambiguated in context. While most of the disambiguation process was quite straightforward, guidelines were put in place to ensure coherence, transparency, and replicability.

Adverbials were included only if they 1) modified a clause or a sentence and 2) were not part of a verbal dependency. Occurrences that were excluded based on 1) are presented in (5) and (6). I followed Dupont's (2019) definition of 'clausal' and 'phrasal', and considered instances with an ellipted verb, like (7), as introducing a clausal element. To determine whether the items were part of verbal dependency, two disambiguation tests were applied (Blanche-Benveniste, 2002; Molinier & Levrier, 2000; Quirk et al., 1985). First, when they could be the focus of a cleft structure, the items were excluded. Secondly, markers that were replaceable by a pronoun or equivalent form were excluded as well. For example, some of the instances of *theoretically* and its French equivalent *théoriquement* were excluded by means of these two tests (see Examples (8) and (9)).

- (5) *Nous avons ensuite le cas de Jelena, dont l'article rapporte les propos au sujet des **apparemment** médiocres performances sexuelles de son ancien partenaire (...)* (FAR, 0091)

‘Then we have the case of Jelena, whose article reports on the **apparently** poor sexual performance of her former partner’

- (6) *We can tell that membership by the second subject in the set denoted by the NP of the first is indeed an important factor by means of the following (**admittedly** awkward) minimal pair.* (KIAP-EN, 0045)
- (7) *Whilst **maybe** not as blatantly obvious as times in the past it would be wrong to think these no longer exist, for evidence we need only to look at the reaction of people when these stereotypes are subverted (...)* (BAWE & MICUSP, 6018f)
- (8) *As with any identifiable aspect of academic writing, much of the research on hedging attempts to define it, **theoretically** and functionally.* (BAWE & MICUSP, LIN.G0.03.1) > *It is **theoretically** that much of the research on hedging is defined*
- (9) *Dans cette partie, les élèves analysent donc **théoriquement** les différentes composantes d'une dissertation (...)* (FAR, 0076) > *(...) les élèves analysent les différentes composantes d'une dissertation **de cette manière***
- 'In this section, pupils **theoretically** analyze the different components of an essay (...)' > '(...) pupils analyze the different components of an essay **in this way**'*

With respect to complement clauses, the literature on English hedging markers tends to be limited to the analysis of 'marker + complement clause introduced by *that/to*' (e.g. Biber, 2006a), while studies on French complement clauses typically focus on segments introduced by *que* (e.g. Boone, 1996; Chevalier & Léard, 1996; Kanté, 2016). These structures are illustrated in Examples (10), (11) and (12). The present study includes a cross-linguistic component, however, and restricting the analysis to these three structures would have led me to overlook key constructions in one language or the other. For

example, as appears from Table 5 below, some English constructions with *that*/*to* (e.g. *consider + to*) correspond to constructions with *de* or with the infinitive in French (e.g. *estimer + INF*). Moreover, in English complement clauses controlled by adjectives and verbs, the marker *that* can be omitted, while this is not the case in French (Biber et al., 1999, see Example (13)). In order to be more cross-linguistically accurate, the present study accounts for a more diverse set of structures: all structures which can be considered as metadiscursive based on the description given in Section 2.2 are included in the analysis. Table 5 illustrates some of those structures (as found in the corpora analyzed).

(10) *There was also a **suggestion** that the bidder may be a financial buyer.*

(Example from Biber et al., 1999: 973)

(11) *If on the other hand liking musicals is mutually **believed** to entail loving opera, then affirmation is implicated.* (KIAP-EN, 0015)

(12) *(...) on peut **penser** que les références à la « beauté », à l'« excellence » ou encore à l'« histoire » seront particulièrement présentes dans le CL (...).* (FAR, 0024)

‘(...) one can **think** that references to “beauty”, “excellence” or “history” will be particularly present in the reading comprehension task (...)’

(13) *We **thought** they concluded this because they counted neoclassical compounds as affixations, too.* (VESPA, UCL0007-LING-01)

Table 5. Examples of complement clause structures included in the present study

STRUCTURE	EXAMPLE (EN)	EXAMPLE (FR)
Marker + <i>de</i> + verb	/	<i>Nous pouvons donc nous avancer à dire que les locuteurs sont susceptibles de produire du wallon à n'importe quel endroit et à n'importe quel moment de leur discours (...) (FAR, 0046)</i> ‘We can therefore say that speakers are likely to produce Walloon at any place and at any time in their discourse (...)’
Marker + subordinate interrogative clause	<i>Uncontroversially, one of the principal goals of historical linguistics is to hypothesize what unattested antecedent forms of a particular language or group of related languages might</i>	/

	<i>have looked like.</i> (KIAP-EN, 0032)	
Marker + omitted <i>that</i> + complement clause	<i>She argued inequalities between the sexes in society were both reflected and constructed by inequalities in language.</i> (BAWE & MICUSP, 6018f)	/
Subject-position clause + Vcopular + marker	<i>However, whether non-verbal deficits are the likely cause or effect of language impairment is unclear with the suggestion that low non-verbal IQ may in fact be found in response to poor comprehension of syntactic structures</i> (Leonard, 2000: 128). (BAWE & MICUSP, LIN.G0.03.1)	/

Marker included in an <i>as comme-</i> clause	<i>As Dudley and Dudley- Evans, 1999 and Thompson (1999) argue, given the increasing numbers of international students undertaking research degrees in English, much more work needs to be done to establish the particular characteristics of the genre they are required to write. (KIAP-EN, 0010)</i>	<i>Comme le suggère Paveau (2011: 42), ce travail n'exclut pas les déclarations émises par les locuteurs non linguistes (...). (FAR, 0044) 'As Paveau (2011: 42) suggests, this work does not exclude statements made by non-linguist speakers (...)</i>
Wh-cleft structure	<i>What seems plausible is that in comprehension children begin on a local syntactic analysis of the input sentence, arriving at an internal structure of the subordinate clause without attention to lexically rooted restrictions on the structure chosen. (KIAP- EN, 0018)</i>	/

The descriptive statistics of raw hedge frequencies after disambiguation are presented in Table 6.

Table 6. Descriptive statistics of raw hedge frequencies after disambiguation in the KIAP-EN, BAWE & MICUSP, VESPA, and FAR

	KIAP-EN	BAWE & MICUSP	VESPA	FAR
<i>ADVERBIALS</i>	Raw freq: 396 Median: 6 IQR: 7.75	Raw freq: 182 Median: 2 IQR: 3	Raw freq: 175 Median: 1 IQR: 2	Raw freq: 243 Median: 3 IQR: 4
<i>ADJECTIVES</i>	Raw freq: 132 Median: 2 IQR: 4	Raw freq: 83 Median: 1 IQR: 2	Raw freq: 50 Median: 0 IQR: 1	Raw freq: 58 Median: 0 IQR: 1
<i>NOUNS</i>	Raw freq: 542 Median: 7.5 IQR: 10.5	Raw freq: 133 Median: 1 IQR: 3	Raw freq: 117 Median: 1 IQR: 2	Raw freq: 140 Median: 1 IQR: 3
<i>VERBS</i>	Raw freq: 2631 Median: 44 IQR: 39.5	Raw freq: 1132 Median: 11 IQR: 12	Raw freq: 1233 Median: 11 IQR: 7	Raw freq: 1200 Median: 13.5 IQR: 15
<i>TOTAL</i>	Raw freq: 3701 Median: 68 IQR: 53	Raw freq: 1530 Median: 16 IQR: 18	Raw freq: 1575 Median: 14 IQR: 11	Raw freq: 1641 Median: 19 IQR: 19

2.5 Statistical methods

The study is restricted to comparisons of frequencies across corpora, and the data examined is not normally distributed (as shown by a Shapiro-Wilk test of normality). Consequently, the choice of a statistical test was narrowed down to monofactorial non-parametric tests (Brezina, 2020; Field et al., 2012). Table 7 provides a summary of the statistical tests used for each type of comparison. The statistical analyses were carried out with the statistical software *R* (R Core Team 2021), and the significance threshold for the tests was set at .05.

Table 7. Statistical tests used to compare hedge and specific marker frequencies across corpora

VARIABLE	OF	NUMBER OF	STATISTICAL TEST	R FUNCTION
INTEREST		CORPORA		
Hedge frequencies		Three corpora	Kruskal-Wallis H test	<code>kruskal.test()</code>
Hedge frequencies		Two corpora	Mann-Whitney U test	<code>wilcox.test()</code>
Specific marker frequencies		Three corpora	Chi-square test of independence	<code>chisq.test()</code>
Specific marker frequencies		Two corpora	Chi-square test of independence	<code>chisq.test()</code>

Comparisons of hedge frequencies were based on the number of occurrences per text normalized per 1,000 words. Comparisons between three corpora were made by means of the Kruskal-Wallis H test, which is based on rank sums. The function ‘`kruskal.test()`’

was computed in *R*. To determine which difference was significant, the *post hoc* Dunn's test was used (`'dunn_test()'` in *R*). Comparisons between two corpora were made by means of the Mann-Whitney U test. The function `'wilcox.test()'` was used to compute the test in *R*. Following Field et al. (2012: 665), the effect size reported for the Mann-Whitney U test is r . In accordance with those two statistical tests, and to account for internal variability in the corpora, the medians, interquartile ranges, and mean ranks are presented as descriptive statistics for comparisons of hedge frequencies in Section 3.

When working with specific markers, it soon became apparent that the texts from the different corpora contain a limited number of items, with many texts containing 0 occurrences of a specific marker. Therefore, the data could not be appropriately handled by the Kruskal-Wallis test or the Mann-Whitney test. As a result, to analyze specific markers I had to fall back on the chi-square test of independence. The `'chisq.test()'` function was performed in *R*. When comparing three groups, Bonferroni corrections of the p-value were computed to control for the Type I error rate (Field et al., 2012: 914)¹³. We follow Gries (2013) and take ϕ as the effect size. The output is a value that ranges from 0 (no effect, according to Cohen's scale) to 1 (very strong effect, according to Cohen's scale). Note, however, that ϕ values reported in LCR studies are typically very small (e.g. Lee et al., 2019; Pan et al., 2016; Plonsky & Oswald, 2014). To interpret those values, one should ideally compare them with the ones reported in similar studies. However, to our knowledge, no study both compares the frequencies of specific markers across corpora and provides an effect size associated with the chi-square test results. The

¹³ The Bonferroni-adjusted p-value is calculated by dividing .05 by the number of comparisons (in this case three), and thus amounts to .017.

outputs of the chi-square tests are included in the supplementary materials (Tables 1.1 to 1.61)¹⁴.

In this study, the use of the terms ‘overuse(d)’, ‘underuse(d)’ or ‘significant(ly)’ means that the p-value is significant (i.e. inferior or equal to .05 or .017, depending on the comparison).

3. Results

This section presents the results of the analysis in terms of hedge frequencies (Section 3.1), lexical preferences (Section 3.2) and diversity of hedge use (Section 3.3). In each of these sub-sections, I proceed in two steps. First, a comparison is drawn between English expert writing, L1 English novice writing, and French EFL learner writing. Second, French EFL learner writing and L1 French novice writing are compared.

3.1 Frequencies

As can be seen from Table 8 and Figure 2, the profile of EFL learners differs from that of English native novices and English experts. EFL learners underuse hedges compared to both groups, with a mean rank of 145.84 in the VESPA compared with 90.86 and 78.01 in the KIAP-EN and in the BAWE & MICUSP respectively. EFL learners thus express the hedging function less frequently than their English native and English expert counterparts. The interquartile range (IQR) is particularly high in the BAWE & MICUSP

¹⁴The supplementary materials can be found here: <https://osf.io/z254u/>.

(6.1), showing a high level of variability across the texts making up the corpus. Note that this variability could be due in part to the fact that the corpus comprises both British and American English texts¹⁵.

Table 8. Descriptive statistics of hedges in English expert, English native novice, and EFL learner writing (based on the number of hedges per text, normalized per 1,000 words)

	ENGLISH EXPERT	ENGLISH NOVICE	EFL LEARNER
	WRITING (KIAP-EN)	WRITING (BAWE & MICUSP)	WRITING (VESPA)
HEDGES	Median: 6.9	Median: 7.5	Median: 4.1
	IQR: 3.425	IQR: 6.1	IQR: 2.85
	Mean rank: 90.86	Mean rank: 78.01	Mean rank: 145.84

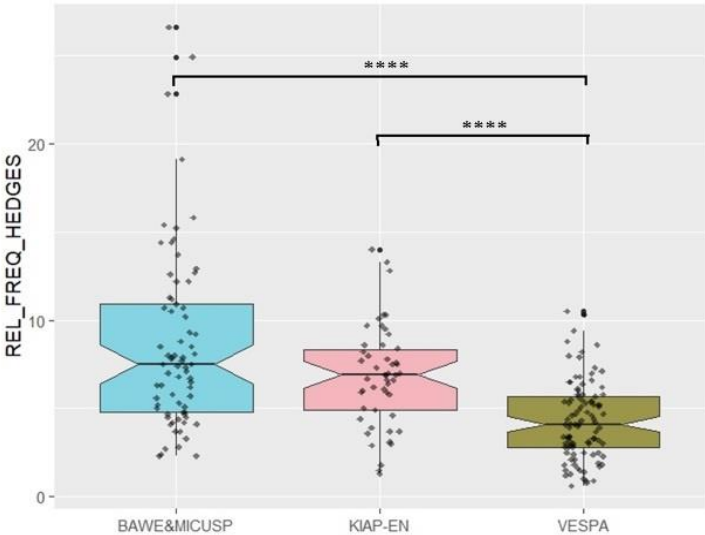


Figure 2. Hedge frequencies in English native novice writing, English expert writing, and

¹⁵ Texts written in an American context might contain fewer hedges than texts written in a British context for instance, due to the tendency among Americans to take on a rather direct and assertive tone (Hall & Hall, 1990). Expanding on the differences between the two varieties is beyond the scope of the present paper, however.

EFL learner writing (Kruskal-Wallis chi-squared = 52.518, df = 2, p-value < .01 ; Dunn post hoc test: p.adj > .05 [BAWE & MICUSP vs. KIAP-EN], p.adj < 0.01 [BAWE & MICUSP vs. VESPA], p.adj < .01 [KIAP-EN vs. VESPA])

Interestingly, a comparison between the VESPA and the FAR indicates that the two corpora show very similar hedge frequencies, with medians of 3.95 and 4.1 hedges per 1,000 words respectively (see Table 9). The Mann-Whitney U test revealed no significant difference between the two groups (see Figure 3).

Table 9. Descriptive statistics of hedges in French native novice and EFL learner writing (based on the number of hedges per text, normalized per 1,000 words)

	FRENCH NOVICE	EFL LEARNER WRITING
	WRITING (FAR)	(VESPA)
<i>HEDGES</i>	Median: 3.95	Median: 4.1
	IQR: 2.275	IQR: 2.85
	Mean rank: 81.33	Mean rank: 86.76

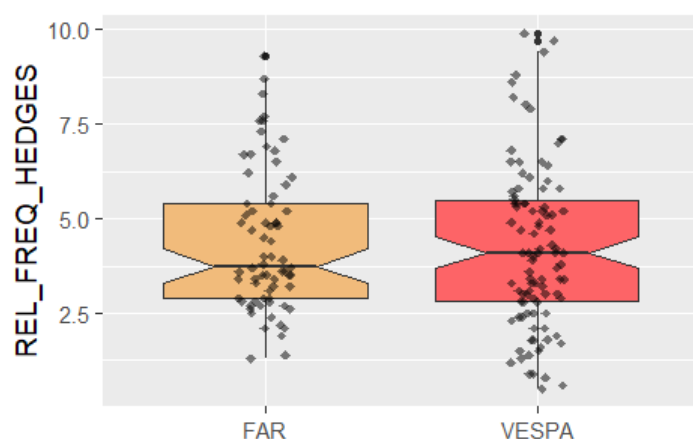


Figure 3. Hedge frequencies in French native novice writing and EFL learner writing (W = 3594, p-value = 0.5977; r: 0.0527748)

3.2 Lexical preferences

Table 10 gives the breakdown of the corpus results per hedge, for the 20 most frequent items in the three English corpora¹⁶. The lexical items are listed in decreasing order of frequency.

The 10 most frequent markers account for the majority of hedges in all the corpora analyzed (from c. 60% of the total number of hedges in the KIAP-EN to c. 73% in the VESPA and FAR), which means that they are a good point of reference to determine whether novices are aware of common hedging devices in academic writing. The 10 most frequent hedges are quite similar in the three populations of English academic writers. In other words, the hedging function is expressed via similar lexical means in the three groups. Among these 10 markers, English native novices share five hedges with experts, and EFL learners share six hedges with experts. The hedges that are common to the three

¹⁶ See supplementary materials (Table 2.1) for the whole table.

populations' 10 most frequent hedges are *seem*, *suggest*, *appear*, and *claim* (v.). The KIAP-EN and the VESPA share the same first ranking marker, namely *seem*. Just like experts and native novices, EFL learners use verbs when expressing doubt, with seven verbs in the KIAP-EN, nine verbs in the BAWE & MICUSP, and eight verbs in the VESPA among the 10 most frequent hedges.

Table 10. 20 most frequent hedges in English expert, English native novice, and EFL learner writing (/100,000 words)

ENGLISH EXPERT			ENGLISH NATIVE		EFL LEARNER WRITING	
WRITING			NOVICE WRITING		(VESPA)	
(KIAP-EN)			(BAWE & MICUSP)			
<i>HEDGE</i>			<i>HEDGE</i>		<i>HEDGE</i>	
<i>REL.</i>			<i>REL.</i>		<i>REL.</i>	
<i>FREQ.</i>			<i>FREQ.</i>		<i>FREQ.</i>	
<i>(PER</i>			<i>(PER</i>		<i>(PER</i>	
<i>100,000</i>			<i>100,000</i>		<i>100,000</i>	
<i>WORDS)</i>			<i>WORDS)</i>		<i>WORDS)</i>	
1	seem	75.2	suggest	132.6	seem	101.0
2	suggest	65.6	seem	101.3	consider	76.7
3	argue	63.7	appear	48.4	appear	31.0
4	assume	54.1	see	43.3	argue	28.5
5	appear	42.2	think	43.3	probably	18.0
6	claim (v.)	34.9	consider	36.4	claim (v.)	14.3
7	claim (n.)	24.9	feel (v.)	36.4	suggest	14.0
8	interpret	24.7	claim (v.)	34.7	think	13.7

9	perhaps	20.3	perhaps	33.0	see	10.2
10	view (n.)	18.5	believe	32.4	not necessarily	7.8
11	propose	18.3	argue	25.6	believe	7.3
12	believe	17.3	assume	23.9	hypothesis	7.3
13	see	16.4	likely (adj.)	23.9	likely (adj.)	7.0
14	assumption	15.1	not necessarily	20.5	opinion	6.5
15	consider	14.9	propose	13.7	assume	5.7
16	not necessarily	14.5	probably	12.0	imply	5.4
17	think	11.1	view (v.)	11.4	maybe	4.8
18	view (v.)	9.4	likely (adv.)	11.4	perhaps	4.3
19	find	8.1	view (n.)	11.4	interpret	4.0
20	likely (adj.)	8.1	possibly	10.2	apparently	3.8
Total		554.2		704.1		370.5

Both L1 English native novices and EFL learners overuse the hedge *seem* compared to experts, and the two novice populations show strikingly similar frequencies of this marker. In addition, both groups underuse the hedges *argue*, *assume*, and *claim* (n.). For *assume* and *claim* (n.), we note a cline in the degree of underuse. *Assume*, for example, displays a frequency of 54.1 occurrences per 100,000 words in the KIAP-EN, 23.9 occurrences per 100,000 words in the BAWE & MICUSP, and 5.7 occurrences per 100,000 words in the VESPA. This – coupled with the learners’ weaker reliance on items like *appear*, *claim* (v.), *view* (n.), *propose*, *assumption* and *view* (v.), which are underused in the learner corpus only – could explain why the VESPA presents a general underuse of hedges but the BAWE & MICUSP does not. The hedge *suggest* ranks first and second

in native novice and in expert writing respectively, but is far rarer in learner writing. Despite the general underuse of hedges, *consider* is particularly frequent in the VESPA, especially as part of the pattern (BE) + *CONSIDER* + *as* (see Example (14)).

(14) According to McArthur (1992: 774), phrasal verbs and multiwords verbs in general are often **considered** as particularly colloquial in comparison to other lexical verbs even though the match between phrasal verbs and simple-word verbs is not straightforward. (VESPA, UCL1213-LING-01)

Some hedges, like *suggest*, *perhaps*, and *believe*, show divergent tendencies across the two novice populations in that they are overused in the BAWE & MICUSP but underused in the VESPA. In the BAWE & MICUSP, we note a strong presence of *see*¹⁷, *think*, *feel*, *likely* (adj.), *likely* (adv.), and *possibly*.

As shown in Table 11, the hedges in the FAR and in the VESPA display similar frequencies¹⁸. The two corpora share the same 2 most frequent formal equivalents, namely *sembler*|*seem* and *considérer*|*consider*. Despite the much lower presence of hedges in French academic writing in comparison with English academic writing (Vold, 2008), *sembler* in the FAR is significantly more frequent than *seem* in the VESPA ($\chi^2 = 6.0133$, $p < .05$). We noted above that *suggest* is used extensively in the KIAP-EN and in the BAWE & MICUSP, but not in the VESPA. This finding is paralleled by the absence of *suggérer* in the 10 most frequent hedges from the FAR. Interestingly, *considérer* and *consider* share very similar frequencies in the two corpora. The two groups also present shared phraseological patterns: (ÊTRE) + *CONSIDÉRER* + *comme* accounts for the majority

¹⁷ With the hedging meaning of ‘perceive’.

¹⁸ See supplementary materials (Table 2.2) for the whole table.

of the occurrences of *considérer* (c. 68%) in the FAR, just like *(BE) + CONSIDER + as* accounts for most of the occurrences of *consider* in the VESPA (c. 77%).

Table 11. 20 most frequent hedges in French native novice writing and EFL learner writing (/100,000 words)

FRENCH NATIVE NOVICE			EFL LEARNER WRITING	
WRITING			(VESPA)	
(FAR)				
	<i>HEDGE</i>	<i>RELATIVE</i>	<i>HEDGE</i>	<i>RELATIVE</i>
		<i>FREQUENCY</i>		<i>FREQUENCY</i>
		<i>(PER 100,000</i>		<i>(PER 100,000</i>
		<i>WORDS)</i>		<i>WORDS)</i>
1	sembler	120.2	seem	101.0
2	considérer	80.2	consider	76.7
3	penser	25.4	appear	31.0
4	hypothèse	21.1	argue	28.5
5	paraître	17.0	probably	18.0
6	peut-être	15.1	claim (v.)	14.3
7	apparaître	14.6	suggest	14.0
8	sans doute	14.3	think	13.7
9	estimer	12.7	see	10.2
10	probablement	10.3	not necessarily	7.8
11	proposer	10.3	believe	7.3
12	supposer	9.2	hypothesis	7.3

13	trouver	7.3	likely (adj.)	7.0
14	certes	6.2	opinion	6.5
15	susceptible	5.9	assume	5.7
16	voir	5.4	imply	5.4
17	éventuellement	4.6	maybe	4.8
18	potentiellement	4.6	perhaps	4.3
19	chance	4.3	interpret	4.0
20	pas forcément	4.3	apparently	3.8
Total		392.9		370.5

Note that, as expected from the literature (see Section 2.5), the effect sizes of the significant differences described in this section are all very small, and therefore hard to interpret¹⁹. It was nonetheless considered important to report them, as they could serve as a point of comparison for future studies.

3.3 Diversity of hedge use

Diversity of hedge use was examined by looking at the cumulative percentages of the most frequent markers. Figure 4 shows the cumulative percentages of the 20 most frequent hedges in English expert, English native novice, and EFL learner writing²⁰. As illustrated, EFL learners exhibit higher percentages than both experts and English native novices. For example, about 50% of hedges are expressed by three markers in the VESPA,

¹⁹ See supplementary materials (Tables 1.1-1.61).

²⁰ A table presenting the specific markers and their corresponding cumulative percentages can be found in the supplementary materials (Table 3.1).

compared to seven markers in the KIAP-EN and in the BAWE & MICUSP. This means that EFL learners display lower diversity of hedge use, in that they tend to resort heavily to the same few items to express uncertainty and caution.

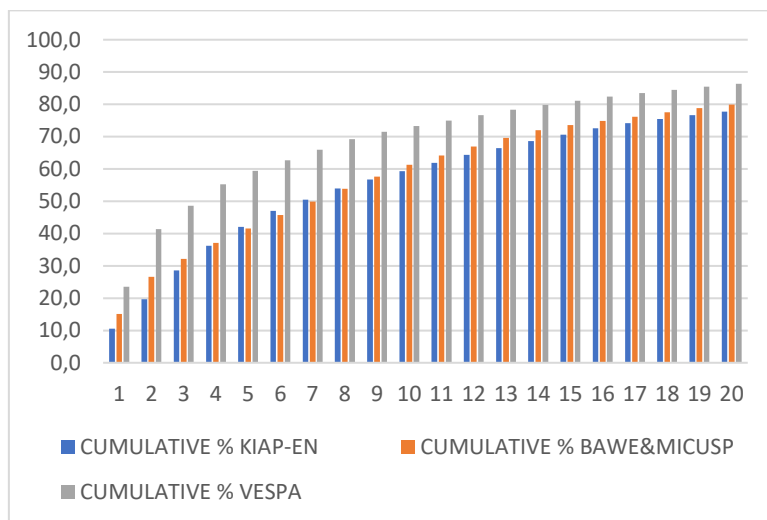


Figure 4. Cumulative percentages of the 20 most frequent hedges in the KIAP-EN, BAWE & MICUSP, and VESPA

As presented in Table 12, the cumulative percentages in the FAR and the VESPA are very similar. For example, the first five markers account for c. 58-59% of all hedges, and the first ten markers account for c. 73% of all hedges in both corpora. In addition to having the same first- and second-ranking markers (namely *sembler|seem* and *considérer|consider*), the first marker (namely *sembler* and *seem*) accounts for c. 24-26% of the total number of tokens, and the first two markers (i.e. *sembler* + *considérer* and *seem* + *consider*) together account for c. 41-44% of the tokens.

Table 12. Cumulative percentages of the 20 most frequent hedges in French native novice writing and in French EFL learner writing

FRENCH NATIVE NOVICE WRITING			EFL LEARNER WRITING	
(FAR)			(VESPA)	
	HEDGE	CUMULATIVE PERCENTAGE	HEDGE	CUMULATIVE PERCENTAGE
1	sembler	26.6%	seem	23.5%
2	considérer	44.3%	consider	41.3%
3	penser	49.9%	appear	48.5%
4	hypothèse	54.5%	argue	55.2%
5	paraître	58.3%	probably	59.4%
6	peut-être	61.6%	claim (v.)	62.7%
7	apparaître	64.9%	suggest	65.9%
8	sans doute	68.0%	think	69.1%
9	estimer	70.8%	see	71.5%
10	probablement	73.1%	necessarily	73.3%
11	proposer	75.4%	believe	75.0%
12	supposer	77.4%	hypothesis	76.7%
13	trouver	79.0%	likely (adj.)	78.3%
14	certes	80.4%	opinion	79.8%
15	susceptible	81.7%	assume	81.2%
16	forcément	82.9%	imply	82.4%
17	voir	84.1%	maybe	83.5%
18	éventuellement	85.1%	perhaps	84.5%

19	nécessairement	86.1%	interpret	85.5%
20	potentiellement	87.1%	apparently	86.3%

4. Discussion

The instances in which LCR studies use a corpus of native novice writing in the same mother tongue as the learners' are rare. Yet the present study makes a strong case for the inclusion of an L1 novice component in the study of learner language. The results yielded by means of a traditional comparison between English expert, L1 English novice, and EFL learner writing, showed that French EFL learners present a profile that differs from L1 English novices and English experts. It confirmed well-documented features of L2 writing, such as a less cautious tone through an underuse of hedges and lower diversity of hedge use (Hinkel, 2005; Hyland & Milton, 1997; Jiang, 2015; Larsson, 2017). These results support the view according to which EFL learners are faced with an additional level of difficulty when compared to (English) L1 novices, due to their non-native status (e.g. Gilquin et al., 2007). On the other hand, the analysis of lexical preferences showed some commonalities across English native novice and EFL learner writing. Both English novice groups exhibit a good mastery of hedging lexical preferences, as shown by the notable overlap between the 10 most frequent hedges and the reliance on verbs to hedge claims across the KIAP-EN, BAWE & MICUSP, and VESPA corpora. It should nonetheless be acknowledged that those findings – especially with regard to hedge frequencies – could be influenced by differences in the levels and text types included in the various corpora. First, although the present paper adopts a binary distinction between

expert and novice writing, the difference between these two language varieties is not as clear-cut in reality, as we are faced with a continuum of writing experience. One could argue that there is not much difference in the level of expertise of an advanced graduate student and an early-career researcher first publishing in a journal, for instance²¹. The BAWE & MICUSP contains successful pieces of writing – therefore possibly placing it on the higher end of the expertise continuum than the two other novice corpora. As a result, the corpus could present more similarities with the KIAP-EN than the VESPA. In addition, as academic writing is less stance-laden than more interpersonal and argumentative registers (e.g. Biber & Finegan, 1989), the more persuasive dimension of some of the texts included in the BAWE & MICUSP (see Section 2.1) could have led those writers to use more hedges than in the other – more expository – corpora. This could explain why VESPA writers underuse hedges, but not BAWE & MICUSP writers. The VESPA writers could also be underusing hedges because their audience consists of reader-evaluators (as using hedges could be perceived as undermining their claims).

The inclusion of a corpus of L1 novice writing permitted better interpretation of the results yielded with the tripartite comparison. It revealed that the features highlighted as typical of learner writing (i.e. a less tentative tone and lower diversity of hedge use) might partly come from the fact that French EFL learners replicate the mechanisms they produce in their L1. This is supported by the evidence yielded from comparing the FAR and the VESPA, as we noted very similar hedge frequencies and cumulative percentages in the two corpora. Therefore, examining L1 novice writing shed light on an important phenomenon at play, namely language-specific register awareness. As mentioned above, English and French expert writing present diverging tendencies in terms of hedge

²¹ Thank you to Reviewer 1 for pointing this out.

frequencies, which necessitates a certain level of adaptation – conscious or subconscious – from the part of writers (Vold, 2008). However, French-speaking students are apparently not aware that they need to adapt their degree of hedging depending on the language of production. They appear to replicate their knowledge of French academic conventions, which tends to result in a style that is less tentative than what is typically expected in terms of English academic standards, for instance.

Analyzing the FAR also highlighted different kinds of transfer. It revealed clear signs of transfer of function (Paquot, 2010; Paquot, 2017): French EFL learners' tendency to avoid expressing uncertainty and caution towards their claims – due to their underuse of hedges – is paralleled in L1 French novice writing. In other words, the tone that emerges from a sentence like (15) (French EFL learner writing), in which the information is presented as a fact rather than as a hedged interpretation from the writer, is also visible in a sentence like (16) (L1 French novice writing). The evidence also indicates the presence of transfer of frequency and transfer of phraseological patterns (Paquot, 2010). A case in point is the verb *consider*, which is overused in French EFL learner writing. The analysis revealed very similar frequencies of *consider* and *considérer* in French EFL learner writing and L1 French novice writing respectively. In addition, both populations extensively use this verb as part of the pattern *(BE) + CONSIDER + as|((ÊTRE) + CONSIDÉRER + comme*. The data therefore shows that writers sometimes over-rely on sequences that are salient in their L1, and transfer them in their L2.

(15) *This phenomenon is due to the reluctance translators have to use or to refer to previous sentences.* (VESPA, UCL1026-LING-01)

(16) *La présence de ces pauses est donc liée au sémantisme des propos tenus et à l'intention du locuteur.* (FAR, 0009)

‘The presence of these pauses is therefore linked to the semantics of the words spoken and to the speaker’s intention’

The present study also made it possible to nuance the results obtained on the basis of a more ‘traditional’ method. Based on the tripartite comparison of between the KIAP-EN, the BAWE & MICUSP, and the VESPA, we saw that both native and non-native novice English writers display a good mastery of hedging lexical preferences. One might conclude that French EFL learners show good adaptation to English academic writing conventions when it comes to selecting the markers to hedge their claims. When comparing the FAR and the VESPA, however, we observed a higher number of similarities between L1 French novice and French EFL learner writing, with considerable lexical overlap across the two groups. For example, the two corpora share the same 2 most frequent formal equivalents (*sembler*|*seem* and *considérer*|*consider*). The good general command of lexical preferences mentioned above might actually not be a sign of mastery of hedging lexical items in EFL learner writing, but rather a mere transfer of the items they use in their L1, to their L2. The result is the same – the texts written by EFL learners do show similar lexical preferences as the ones written by L1 English novices – but the interpretation is nuanced: this feature could be the result of favorable cross-linguistic circumstances, thanks to the similarities between the learners’ L1 and L2.

5. Conclusion

This paper aimed to answer the following research question: “To what extent does adding a corpus of novice writing in the L1 of the learners complement the LCR methodological apparatus and contribute to a better understanding of the complex interplay of factors that influence learner language?”. To do so, a corpus-based study was carried out on the use of hedges in French EFL learner academic writing. It combined a more traditional tripartite approach between English expert writing, English native novice writing and French EFL learner writing, with the use of a L1 novice corpus of French academic writing. It was found that while EFL learner writing and L1 English novice writing diverge in some aspects and converge in others, learners tend to replicate their own L1 French hedge usage in their L2 English written productions. The findings speak to the importance of including an L1 novice component to the current methodologies employed in LCR, as such inclusion helps to provide a better and more nuanced interpretation of the results obtained by means of a more traditional tripartite approach. For instance, various kinds of L1 transfer were found to impact learner use of hedges, namely transfer of function, transfer of frequency, and transfer of phraseological patterns.

The findings reported in this paper should also be considered in light of their limitations. The conclusions on L1 and L2 writing by French-speaking students are based on written productions from two comparable yet separate sets of writers. The results yielded from the study provide strong evidence that analyzing L1 and L2 data from different individuals already gives valuable insights into learner language specificities. Given the importance of individual variation (Wulff & Gries, 2021), however, analyzing

a corpus of L1-L2 writing by the same individuals could help broaden our understanding of those specificities even more.

On the basis of the findings presented in this article, it is argued that the field of LCR lacks L1 and L2 novice corpora that follow the same corpus design. Not only is it desirable for the field to compile L1 novice corpora that are comparable to L2 learner corpora, but – as explained above – there is also a need to collect data in L1 and L2 by the same novice writers. The field is nonetheless moving in this direction, with corpora such as the *Aachen Corpus of Academic Writing* (ACAW), which contains L1 German and L2 English writing by the same individuals (Ströbel et al., 2020).

References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins.
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92.
- Ädel, A., & Römer, U. (2012). Research on advanced student writing across disciplines and levels. Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics*, 17(1), 3–34.
- Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116.
- Biber, D. (2006b). *University language. A corpus-based study of spoken and written registers*. John Benjamins.

- Biber, D., Egbert, J., & Zhang, M. (2018). Lexis and grammar as complementary discourse systems for expressing stance and evaluation. In M. Gómez González & J. L. Mackenzie (Eds.), *The Construction of Discourse as Verbal Interaction* (pp.201–226). John Benjamins.
- Biber, D., & Finegan, E. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text*, 9(1), 93–124.
- Biber, D., Johansson, S., Leech, G. Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. Longman.
- Biber, D., Reppen, R., Staples, S., & Egbert, J. (2020). Exploring the longitudinal development of grammatical complexity in the disciplinary writing of L2-English university students. *International Journal of Learner Corpus Research*, 6(1), 38-71.
- Blanche-Benveniste, C. (2002). Phrase et construction verbale. *VERBUM*, XXIV(1), 7–22.
- Boone, A. (1996). Les complétives et la modalisation. In C. Muller (Ed.), *Dépendance et intégration syntaxique: Subordination, coordination, connexion* (pp. 45–51). Niemeyer.
- Brezina, V. (2020). Classical monofactorial (parametric and non-parametric) tests. In M. Paquot & S. Th. Gries (Eds.). *A Practical Handbook of Corpus Linguistics* (pp. 473–503). Springer.
- Çandarlı, D., Bayyurt, Y., & Martı, L. (2015). Authorial presence in L1 and L2 novice academic writing: Cross-linguistic and cross-cultural perspectives. *Journal of English for Academic Purposes*, 20, 192–202.

- Chen, H.-I. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Arizona Working Papers in SLA & Teaching*, 17, 27–51.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Chen, Z. (2012). Expression of epistemic stance in EFL Chinese university students' writing. *English Language Teaching*, 5(10), 173–179.
- Chevalier, G., & Léard, J.-M. (1996). La subordination nominale: classes, sous-classes et types sémantiques. In C. Muller (Ed.), *Dépendance et intégration syntaxique: Subordination, coordination, connexion* (pp. 52–65). Niemeyer.
- Crosthwaite, P., Cheung, L., & Jiang, K. (2017). Writing with attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes*, 46, 107–123.
- Dirdal, H. (2022). Cross-linguistic influence in the acquisition of relative clauses by Norwegian learners of English. *Nordic Journal of Language Teaching and Learning*, 10(2), 327–351.
- Dupont, M. (2019). *Conjunctive Markers of Contrast in English and French. From Syntax to Lexis and Discourse* (Unpublished doctoral dissertation). Université catholique de Louvain.
- Ebeling, S. O., & Hasselgård, H. (2015). Learners' and native speakers' use of recurrent word-combinations across disciplines. *Bergen Language and Linguistics Studies (BeLLS)*, 6, 87–106.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE.
- Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic Voices*. John Benjamins.

- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: Detection, explanation, evaluation. In G. Gilquin, S. Papp, & M. B. Díez-Bedmar (Eds.), *Linking Up Contrastive and Learner Corpus Research* (pp. 3–33). Rodopi.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: the missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335.
- Granger, S., & Paquot, M. (2010). The Louvain EAP Dictionary (LEAD). In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the XIV EURALEX International Congress* (pp. 321–326). Fryske Akademy.
- Gray, B., & Biber, D. (2014). Stance markers. In K. Ajimer & C. Rühlemann (Eds.), *Corpus Pragmatics: A Handbook* (pp. 219–248). Cambridge University Press.
- Gries, S. Th. (2013). *Statistics for Linguistics with R. A practical introduction* (2nd edition). De Gruyter Mouton.
- Hall, E. T., & Hall, R. H. (1990). *Understanding cultural differences. Germans, French and Americans*. Intercultural Press.
- Heuboeck, A., Holmes, J., & Nesi, H. (2008). *The BAWE corpus manual*. <http://www.reading.ac.uk/internal/applying/bawe/BAWE.documentation.pdf>
- Hinkel, E. (2005). Hedging, inflating and persuading in L2 academic writing. *Applied Language Learning*, 15(1&2), 29–53.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183–205.
- Hyland, K., & Sancho Guinda, C. (Eds.). (2012). *Stance and voice in written academic genres*. Palgrave Macmillan.

- Hyland, K., & Tse, P. (2005). Hooking the reader: a corpus study of evaluative *that* in abstracts. *English for Specific Purposes*, 24(2), 123–139.
- Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, 50(2), 245–309.
- Jiang, F. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes*, 20, 90–102.
- Kanté, I. (2010). Mood and modality in finite noun complement clauses. A French-English contrastive study. *International Journal of Corpus Linguistics*, 15(2), 267–290.
- Kanté, I. (2016). Contraintes syntaxiques et modales des noms recteurs dans la complétive finie en français. In F. Neveu, G. Bergounioux, M.H. Côté, J.-M. Fournier, L. Hriba, & S. Prévost (Eds.), *Actes du CMLF 2016 – 5e Congrès mondial de Linguistique française*. SHS Web of Conferences.
- Kanté, I. (2019). Le nom recteur et l'expression de la modalité en anglais et en français. *Épilogos*, 6, 243–164.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7–36.
- Larsson, T. (2016). *The introductory it pattern in academic writing by non-native-speaker students, native-speaker students and published writers: A corpus-based study* (Unpublished doctoral dissertation). Uppsala University.
- Larsson, T. (2017). A functional classification of the introductory-*it* pattern: Investigating academic writing by non-native speaker and native speaker students. *English for Specific Purposes*, 48, 57–70.

- Larsson, T. (2019). Grammatical stance marking across registers: Revisiting the formal-informal dichotomy. *Register Studies*, 1(2), 243–268.
- Larsson, T., Reppen, R., & Dixon, T. (2022). A phraseological study of highlighting strategies in novice and expert writing. *Journal of English for Academic Purposes*, 60, 1–10.
- Lee, J. J., Bychkovska, T., & Maxwell, J. D. (2019). Breaking the rules? A corpus-based comparison of informal features in L1 and L2 undergraduate student writing. *System*, 80, 143–153.
- Lee, J.J., & Deakin, L. (2016). Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing*, 33, 21–34.
- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on Computer* (pp. xiv–xx). Longman.
- Lorenz, G. (1999). *Adjective intensification - Learners versus native speakers. A corpus study of argumentative writing*. Rodopi.
- Molinier, C., & Levrier, F. (2000). *Grammaire des adverbes. Description des formes en –ment*. Droz.
- Neff, J., Dafouz, E., Díez, M., Martínez, M., Prieto, R., & Rica-Peromingo, J.-P. (2003). Evidentiality and the construction of writer stance in native and non-native texts. In J. Hladký (Ed.), *Language and Function: To the Memory of Jan Fibras* (pp. 223–235). John Benjamins.
- Nesi, H. (2008). BAWE: An Introduction to a New Resource. In A. Frankenberg-Garcia, R. Tawfiq, M. do Rosário Braga da Cruz, R. Carvalho, C. Direito, & D. Santos-

- Rosa (Eds.), *Proceedings of the 8th Teaching and Language Conference* (pp. 239–246). Instituto Superior de Linguas e Administracao.
- O'Donnell, M. B., & Römer, U. (2012). From student hard drive to web corpus (part 2): the annotation and online distribution of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 7(1), 1–18.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60–71.
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis*. Continuum.
- Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research*, 33(1), 13–32.
- Paquot, M., Larsson, T., Hasselgård, H., Ebeling, S. O., De Meyere, D., Valentin, L., Laso, N. J., Verdaguer, I., & van Vuuren, S. (2022). The Varieties of English for Specific Purposes dAtabase (VESPA): Towards a multi-L1 and multi-register learner corpus of disciplinary writing. *Research in Corpus Linguistics*, 10(2), 1–15.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Römer, U. (2009). English in academia: Does nativeness matter? *Anglistik: International Journal of English Studies*, 20(2), 89–100.

- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159–177.
- Rørvik, S. (2021). Cross-disciplinary and cross-linguistic uses of tensed verb phrases in the methods sections of master's theses. In A. Čermáková, T. Egan, H. Hasselgård, & S. Rørvik (Eds.), *Times in Languages, Languages in Time*. John Benjamins.
- Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing: Applying a corpus-based study to language teaching*. John Benjamins.
- Schweinberger, M. (2020). A corpus-based analysis of differences in the use of *very* for adjective amplification among native speakers and learners of English. *International Journal of Learner Corpus Research*, 6(2), 163–192.
- Scott, M., & Tribble, C. (2006). *Textual patterns. Key words and corpus analysis in language education*. John Benjamins.
- Shin, Y. K. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundle in college students' essays. *Journal of English for Academic Purposes*, 40, 1–14.
- Sian, A., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71–83.
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexicogrammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, 32, 17–35.
- Ströbel, M., Kerz, E., & Wiechmann, D. (2020). The relationship between first and second language writing: Investigating the effects of first language complexity on

- second language complexity in advanced stages of learning. *Language Learning*, 70(3), 732–767.
- Su, H., Zhang, Y., & Chau M.H. (2022). Exemplification in Chinese English-major MA students’ and expert writers’ academic writing: A local grammar based investigation. *Journal of English for Academic Purposes*, 58, 1–12.
- Taqvist, M. K. (2018). “A wise decision”: Pre-modification of discourse-organizing nouns in L2 writing. *Journal of Second Language Writing*, 41, 14–26.
- Vold, E. T. (2008). *Modalité épistémique et discours scientifique. Une étude contrastive des modalisateurs épistémiques dans des articles de recherche français, norvégiens et anglais, en linguistique et médecine* (Unpublished doctoral dissertation). University of Bergen.
- Wang, J., & Jiang, F. (2018). Epistemic stance and authorial presence in scientific research writing. Hedges, boosters and self-mentions across disciplines and writer groups. In P. Mur-Dueñas & J. Šinkūnienė (Eds.), *Intercultural perspectives on Research Writing* (pp. 195–216). John Benjamins.
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC*, 42(2), 155–166.
- Wulff, S., & Gries, S. Th. (2021). Exploring individual variation in learner corpus research: Methodological suggestions. In B. S. W. Le Bruyn & M. Paquot (Eds.), *Learner Corpus Research Meets Second Language Acquisition* (pp. 191-213). Cambridge University Press.
- Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics*, 50 (1), 23–36.

Corresponding Author

Pauline Jadoulle

Université catholique de Louvain

FIAL

Place Cardinal Mercier 31/L3.03.33

1348 Louvain-la-Neuve

Belgium

pauline.jadoulle@uclouvain.be

Publication history

Date received: 9 December 2022

Date accepted: 6 November 2023