

An investigation of stance in French native novice academic writing

Pauline Jadoulle

UCLouvain

In the past decades, studies in the field of ‘littéracies universitaires’ (‘academic literacies’) have highlighted a number of evaluative features typical of French novice academic writing, such as an assertive and personal tone (e.g. Donahue 2002, Rinck 2011). However, such studies are largely qualitative, usually focus on a limited number of texts, and are typically based on theoretically-grounded concepts such as ‘posture’ and ‘image du scripteur’ (‘writer’s image’). The present study aims to fill this gap by conducting a quantitative corpus-based investigation of French novice writing, and using a larger set of data. The analyses are carried out by comparing a corpus of French novice academic writing with a corpus of French expert academic writing. The focus is placed on stance, and more precisely on epistemic and attitudinal adverbials and complement clauses. Results help to nuance previous findings on French novice writing. On the one hand, French novices do not necessarily take on a more assertive tone than French experts. In fact, they were found to be more tentative than their expert counterparts as they tend to overuse hedges like *sembler* and *peut-être*. On the other hand, novices were shown to adopt a more attitude-laden tone than experts, which is in line with findings highlighted by more qualitative studies.

1. Introduction

In the past decades, research in the field of ‘littéracies universitaires’ (‘academic literacies’) has highlighted a number of features typical of French novice academic writing, especially with respect to concepts linked to evaluative meanings such as ‘posture’, ‘ethos’, or ‘image du scripteur’ (‘writer’s image’) (e.g. Donahue 2002, Colin & Dolignier 2019, Rinck 2011). Although there is considerable definitional uncertainty and blurriness surrounding these concepts, it can be said that the latter tend to refer to how writers present themselves in their texts, and how they relate to various aspects such as knowledge, language, and writing (Delcambre & Reuter 2002). For instance, Rinck (2011: 80) defines students’ posture in their writing as “an effect of the teacher’s reading of it”, and Fløttum & Vold (2010: 42) define ‘ethos’ as “the self-image that the writer projects and that is produced by discourse”¹. One feature commonly associated with French novice writing is the adoption of an assertive tone; French novices are said to present their claims as self-evident, as general truths, and sometimes even as prescriptions (e.g. Deronne 2011, Delarue-Breton 2014). Those writers are also said to adopt a style that is too personal and attitude-laden (that is, filled with attitudes and value judgments) compared to academic writing standards, which can negatively impact the reader-evaluator’s assessment of their texts (e.g. Rinck & Pouvreau 2009, Reuter 2004).

These findings are mainly based on qualitative and small-scale analyses of internship reports, however, and are generally theoretically-grounded (researchers allocate much space to the discussion of theoretical concepts such as ‘la créativité’ (‘creativity’) (Delarue-Breton 2014) or ‘la nature dialogique’ (‘dialogical nature’) (Donahue 2002)). For example, Donahue (2002) and Delcambre & Laborde-Milaa (2002) analyze three student papers and dissertation introductions respectively, Delarue-Breton (2014) focuses on 12

¹ All translations are mine.

documents written by six students, and Rinck & Pouvreau (2009) examine 29 first sections of internship reports². These researchers' methodology apparently consists in conducting a close reading of the texts under analysis. There is arguably a need for more corpus-based approaches involving larger-scale quantitative research, as this would arguably help to understand whether the findings on French novice academic writing can be generalized.

To fill this gap, the present study applies a corpus-based quantitative analysis of epistemic and attitudinal stance features to corpora of French academic writing. 'Stance' is usually analyzed in English academic writing, and is defined as the linguistic ways writers express their assessment and commitment, and attitudes and value judgments towards the information presented (e.g. Biber et al. 1999). Stance thus overlaps semantically with the previously mentioned concepts of 'posture', 'ethos' and 'image d'auteur' ('writer's image'), but, contrary to these concepts, it is often studied through quantitative approaches of a wide range of lexico-grammatical features (e.g. Biber 2006). Investigating stance features will therefore make it possible to quantitatively revisit the features commonly associated with French novice writing.

2. Data

To get a better understanding of stance use in French novice academic writing, a comparison is drawn between a corpus of French novice writing and a corpus of French expert academic writing.

The corpus of French novice writing used in the present study is the *French Academic wRiting* (FAR) corpus. It consists of 70 academic papers written by French-speaking students in the framework of their linguistics courses at UCLouvain (with a total of

² None of these authors mentions the number of words contained in the corpora analyzed.

370,353 words). 19 papers were written by third-year BA students and 51 papers were written by MA students.

The second corpus analyzed is the KIAP-FR-LING (henceforth KIAP-FR), which is a sample from the *Kulturell Identitet i Akademisk Prosa: nasjonal versus disiplinavhengig* (Fløttum et al. 2006). It comprises 50 research articles written in French in the discipline of linguistics, and consists of 303,342 words. The texts are mainly expository and “primarily report and discuss empirical (...) or theoretical research that has been done by the authors themselves” (Fløttum et al. 2006: 8). The authors of these texts are professional academics whose papers were published in peer-reviewed journals³.

3. Defining stance

Stance involves two main categories, namely epistemic meanings and attitudinal meanings. Epistemic stance conveys the ways writers comment on the “status of information in a proposition” (Biber et al. 1999: 972), and more specifically writers’ assessment of and level of commitment towards the degree of certainty and reliability of a proposition (Hyland 2005, Biber et al. 1999). Epistemic stance is subdivided into hedges and boosters. While hedges (e.g. *peut-être*, *supposition*) allow writers to express caution, uncertainty and distance towards the information presented, boosters (e.g. *évidemment*, *certain*) have the opposite effect in that they help writers show certainty, involvement and confidence towards the propositions they put forward. Attitudinal stance indicates “writer’s affective, rather than epistemic, attitude to propositions, conveying

³ 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).

surprise, agreement, importance, frustration, and so on, rather than commitment” (Hyland 2005: 180). Common attitudinal stance meanings include ‘importance’ (e.g. *important*), ‘agreement’ (e.g. *admettre*), ‘happiness’ (e.g. *heureusement*), ‘interest’ (e.g. *intéressant*), ‘difficulty’ (e.g. *facile*), ‘surprise’ (e.g. *surprenant*), ‘success’ (e.g. *échouer*), and ‘judgment’ (e.g. *juger*).

This study focuses on metadiscursive stance markers: it only considers markers that, in the same sentence, clearly qualify propositional content by means of two overt components, namely a stance marker, and a proposition framed by this marker (Gray & Biber 2014). For example, in sentences (1) and (2), the markers *suggérer* and *évident* frame the proposition included in the *que*-clause (“que ces ‘actions’ (...) permettent bien plus que ce qu’on peut faire d’homologue avec son corps ou des algorithmes” and “que ces deux prépositions ont deux lectures” respectively). By contrast, the same markers in sentences (3) and (4) are excluded from the study, as they do not frame clausal content.

- (1) *En même temps, il est **évident** que ces “actions” (...) permettent bien plus que ce qu’on peut faire d’homologue avec son corps ou avec des algorithmes.* (KIAP-FR, 0011)
- (2) *Dans cette liste de prépositions, deux synonymes définitionnels à cause de *pour* au sujet de *et* à l’occasion de *pour* à propos de *surprennent* et **suggèrent** que ces deux prépositions ont deux lectures.* (KIAP-FR, 0009)
- (3) *D’où les risques **évidents** d’oubli, de légèreté, de banalité, d’“hors -sujet” (...)* (KIAP-FR, 0011)
- (4) *(...) nous **suggérons** ici une méthodologie permettant de faire une étude sur le verlan aujourd’hui (...)* (FAR, 0019)

Three main grammatical features can be said to convey stance metadiscursively, namely adverbials, complement clause constructions controlled by adjectives, nouns, and verbs, and modal verbs (Biber et al. 2018)⁴. This study focuses on adverbials and adjectives, nouns, and verbs controlling complement clauses.

4. Methodology

This section gives an overview of the methodology adopted to identify and extract markers, and disambiguate them in context. It also presents the statistical tests conducted when comparing the two corpora. As will be evident, the methodology adopted is distinct from conventional frameworks adopted in the French tradition, where researchers typically conduct qualitative analyses of student texts (see Section 1). The present study provides an arguably structured and replicable methodology for the analysis of evaluative meanings in French academic writing, and makes it possible to systematically and quantitatively compare student productions to expert ones. This approach will hopefully provide new insights into the ways novice writers convey their assessment, commitment, attitudes and value judgments in academic writing.

4.1 Marker identification and extraction

Adverbials were identified by extracting adverbs and n-grams (i.e. continuous sequences of ‘n’ words) from the corpora, and adjectives, nouns and verbs were identified by extracting common stance patterns such as ADJ + *de* and VERB + *que*. This section describes the identification process in more detail. Note that the selection of stance markers was made in close comparison and consultation with

⁴ ‘Stance noun + prepositional phrase’ constructions are also used to express stance (e.g. *the possibility of a death wish*; Biber et al. 1999: 970).

pre-existing lists of English markers provided in stance studies, like Hyland (2005a), Aull (2015), Biber (2006), and Pho (2013).

4.1.1 Adverbials

The category of adverbials includes single-word adverbs (e.g. *probablement*) and multi-word adverbial expressions (e.g. *sans doute*). To identify single-word adverbials, I used Sketch Engine (Kilgarriif et al. 2014) to extract the tag [tag=".*R"] (i.e. all items tagged as adverbs) from the corpora. I then subjected the resulting items to two criteria so as to decide whether or not to include them in the study: (a) the item fits the semantic definition of stance provided above, and (b) the item is used metadiscursively. This was verified by considering the items in context. Multi-word adverbial expressions were found by extracting 2- and 3-grams (minimum frequency = 3) from the corpora. Out of 51,855 n-grams, 6 items were kept.

The single- and multi-word adverbials yielded with this method are presented in Table 1.

Table 1. Adverbials under study⁵

Boosters		Hedges	Attitudinal markers
absolument	irréductiblement	apparemment	à juste titre
assurément	inéluctablement	censément	crucialement
bien entendu	inévitablement	certes	curieusement
bien sûr	manifestement	éventuellement	étonnamment
certainement	naturellement	peut-être	étrangement
clairement	nécessairement	possiblement	heureusement
définitivement	obligatoirement	potentiellement	idéalement
évidemment	ostensiblement	prétendûment	ironiquement
fatalement	oui	probablement	légitimement
forcément	résolument	sans doute	logiquement
immanquablement	sans aucun doute	supposément	malheureusement
impérativement	sûrement	théoriquement	paradoxalement
impérieusement		visiblement	raisonnablement
incontestablement		vraisemblablement	rien d'étonnant
indéniablement			significativement
indubitablement			valablement

⁵ In the table (as well as in Tables 3, 4 and 5 below), the items are associated with their main semantic category. Note however that some markers can belong to various semantic categories depending on the context. For example, when preceded by the negative marker *pas*, *nécessairement* functions as a hedge rather than as a booster (e.g. *Or, cette ressource n'est pas **nécessairement** facilement accessible à tout le monde*; FAR, 0031).

4.1.2 Adjectives, nouns, and verbs

Figure 1 summarizes the main steps undertaken to find controlling adjectives, nouns, and verbs.

- | |
|---|
| <ol style="list-style-type: none">1. Selection of prominent adjectival, nominal and verbal stance patterns (e.g. ADJ + <i>de</i>, VERB + <i>que</i>)2. Extraction of these patterns in the corpora under study3. Selection of controlling words based on two criteria (i.e. stance meaning and metadiscursive status)4. Extraction of these controlling words5. Disambiguation in context (see Section 4.2) |
|---|

Figure 1. Main steps to find complement clauses

First, prominent stance patterns needed to be selected (step 1 in Figure 1). To find those patterns, 2- to 6-grams including common stance adjectives, nouns and verbs⁶ (minimum frequency = 5) were extracted from a large corpus of French academic writing, namely the LOCRA-FR⁷. The LOCRA-FR contains 313 French articles from top-ranked journals in the Humanities, written by expert writers. It contains c.2 million words and it is made up of articles from five disciplines: anthropology, education, political science, psychology and sociology⁸. As shown in Table 2, the stance patterns

⁶ 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 lists of 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>, last accessed 20 September 2022).

⁷ <https://uclouvain.be/en/research-institutes/ilc/cccl/locra.html> (last accessed 23 March 2023).

⁸ 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

found by means of this method were the following: ADJ + *que|qu' |de|d' |à*, NOUN + *que|qu' |de|d' |selon|à* and VERB + *que|qu' |de|d' |à|INF*. Some examples of these patterns in context are illustrated in (5), (6) and (7).

- (5) *En effet, et c'est sans doute le plus **intéressant** à analyser, la contribution turque au renouveau de l'enseignement de l'islam est surtout l'œuvre de nombreux mouvements privés.* (LOCRA-FR, SCP008-05)
- (6) *Même dans ce cas, il reste une **possibilité** que la variable à quatre niveaux (non-croyants, croyants non pratiquants, croyants pratiquants et religieux engagés à vie) considérée dans ces études ne reflète pas parfaitement le niveau réel de religiosité des participants.* (LOCRA-FR, PSY0006-03)
- (7) *Par contre, il **importe** d'être vigilant à ce changement, car ceci pourrait amener les étudiants à moins s'investir dans les premières étapes de la démarche et à laisser le tuteur reprendre les rênes, devenant ainsi plus passifs et réintégrant le paradigme de l'enseignement traditionnel.* (LOCRA-FR, EDU001-03)

Table 2. Extraction of n-grams with common stance markers from the LOCRA-FR to find prominent stance patterns

Marker	Grammatical category	Semantic category	Examples of stance n-grams extracted	Stance patterns found
<i>ÉVIDENT</i>	adjective	booster	il est évident que évident qu'	évident que évident qu'

needed to complete this step, so as not to miss important stance patterns (which means that I could not use a corpus like the KIAP-FR or the FAR for instance).

An investigation of stance in French native novice academic writing 11

<i>SÛR</i>	adjective	booster	sûr que sûr de	sûr que sûr de
<i>POSSIBLE</i>	adjective	hedge	est possible que possible qu'	possible que possible qu'
<i>PROBABLE</i>	adjective	hedge	probable que probable de probable qu'	probable que probable de probable qu'
<i>IMPORTANT</i>	adjective	attitudinal	important que important de noter il est important d'	important de important d' important que
<i>INTÉRESSANT</i>	adjective	attitudinal	intéressant d' intéressant à il est intéressant de	intéressant de intéressant d' intéressant à
<i>AFFIRMATION</i>	noun	booster	affirmation de	affirmation de
<i>CERTITUDE</i>	noun	booster	/	/
<i>HYPOTHÈSE</i>	noun	hedge	l'hypothèse que l'hypothèse selon laquelle l'hypothèse qu'un	hypothèse que hypothèse selon (laquelle) hypothèse qu'
<i>POSSIBILITÉ</i>	noun	hedge	possibilité que	possibilité que
<i>DIFFICULTÉ</i>	noun	attitudinal	difficulté de difficulté à la difficulté d'	difficulté de difficulté à difficulté d'
<i>OPINION</i>	noun	attitudinal	/	/
<i>AFFIRMER</i>	verb	booster	affirmer que affirme qu'	affirmer que affirmer qu'
<i>DÉMONTRER</i>	verb	booster	démontrer qu' démontrer que les	démontrer qu' démontrer que
<i>SUGGÉRER</i>	verb	hedge	suggère que suggère qu' suggéré de	suggérer que suggérer qu' suggérer de

<i>ESTIMER</i>	verb	hedge	estimer que estime qu' estiment être	estimer que estimer qu' estimer + INF
<i>ARRIVER</i>	verb	attitudinal	arrive à	arriver à
<i>IMPORTER</i>	verb	attitudinal	importe de importe d' importe que	importer de importer d' importer que

Those patterns were then extracted from the KIAP-FR and the FAR in Sketch Engine (step 2 in Figure 1), with the following queries:

- [tag="A.*"]⁹ [word="de|d|que|qu|à"] (for adjectives)
- [tag="N.*"]¹⁰ [word="que|qu|de|d|selon|à"] (for nouns)
- [tag="V.*"]¹¹ [word="que|qu|de|d|à"] and [tag="V.*" & lemma!="avoir|être|pouvoir|devoir|faire|sembler|paraître|fail
lir"¹²|falloir"] [tag="V.*"]¹³ (for verbs)

These searches yielded about 1,000 adjectives, 2,700 nouns, and 1,000 verbs. Those items were subjected to the same two criteria as the ones used to select adverbials (step 3 in Figure 1). The selection process resulted in a total of 87 adjectives, 60 nouns, and 77 verbs (see Tables 3, 4 and 5 below), which were then extracted from the corpora (step 4 in Figure 1).

⁹ All items tagged as adjectives.

¹⁰ All items tagged as nouns.

¹¹ All items tagged as verbs.

¹² The program mistakenly considered *faillir* as the infinitive of *faut* and *faillie*.

¹³ This corresponds to all verbs except the most frequent (semi-)modals – which were creating a lot of noise – followed by another verb.

Table 3. Controlling adjectives under study

Boosters	Hedges	Attitudinal markers			
certain	admissible	absurde	essentiel	loisible	surpris
clair	contestable	abusif	étonnant	malaisé	utile
connu	enclin	agaçant	étrange	mieux	
convaincant	improbable	aisé	facile	normal	
convaincu	possible	amusant	faux	opportun	
évident	probable	anodin	fondamental	pertinent	
impératif	susceptible	ardu	fortuit	préféré	
incontestable	vraisemblable	banal	frappant	prétentieux	
indéniable		complexe	hasardeux	primordial	
indispensable		compliqué	heureux	productif	
justifié		compréhensible	important	raisonnable	
manifeste		condamné	inintéressant	regrettable	
naturel		conscient	injuste	remarquable	
nécessaire		contraignant	injustifié	révélateur	
obligatoire		crucial	intéressant	risqué	
sûr		curieux	inutile	significatif	
vrai		délicat	judicieux	simple	
		difficile	juste	soucieux	
		éclairant	légitime	souhaitable	
		erroné	logique	surprenant	

Table 4. Controlling nouns under study

Boosters	Hedges	Attitudinal markers		
affirmation	argument	avantage	légitimité	tentation
besoin	avis	complexité	lucidité	tort
certitude	chance	conscience	mérite	utilité
conclusion	doute	crainte	occasion	
constat	hypothèse	défaut	opportunité	
constatation	impression	défi	particularité	
devoir	intuition	difficulté	pertinence	
impératif	possibilité	erreur	peur	
injonction	postulat	espérance	préoccupation	
nécessité	présomption	espoir	prétention	
obligation	sentiment	exploit	problème	
preuve	supposition	facilité	regret	
	thèse	illusion	revendication	
		importance	risque	
		intérêt	sagesse	
		jugement	stratégie	

Table 5. Controlling verbs under study

Boosters		Hedges		Attitudinal markers		
affirmer	NEG + nier	apparaître	ressentir	accepter	juger	risquer
assurer	persuader	argumenter	sembler	admettre	manquer	valoir
attester	prouver	conjecturer	sentir	apprécier	menacer	veiller
conclure	savoir	considérer	sous-entendre	avouer	mériter	
confirmer	témoigner	croire	soutenir	concéder	objecter	
constater	trouver	décréter	suggérer	se contenter	opposer	
convaincre	voir	douter	supposer	convenir	oser	
déclarer		estimer		craindre	oublier	
démontrer		imaginer		défendre	préférer	
établir		inférer		échouer	prétendre	
NEG + hésiter		paraître		espérer	réaliser	
justifier		penser		étonner	reconnaître	
manifester		postuler		exiger	regretter	
montrer		présumer		importer	rétorquer	
nécessiter		proposer		intéresser	revendiquer	

4.2 Data disambiguation

Once the items presented in Tables 1, 3, 4 and 5 were extracted, I disambiguated the occurrences in context to determine if they qualified as instances of metadiscursive stance markers. The results outlined in Section 5 reflect the findings obtained post-disambiguation.

Adverbials were only included if they (a) modified a clause or a sentence and (b) were not part of verbal dependency (Blanche-Benveniste 2002, Molinier & Levrier 2000, Quirk et al. 1985). Examples of occurrences excluded based on (a) and (b) are presented in (8) and (9) respectively.

- (8) (...) *Pauline reprend de manière quasi identique certaines de ces phrases répétées sans doute déjà plusieurs fois (et de manière **certainement** inconsciente).* (FAR, 0036)

- (9) *De cette manière, on peut comprendre plus **assurément** pourquoi les chercheurs ne peuvent se détacher de l'un ou de l'autre champ (...)* (FAR, 0054)

With respect to ‘adjectives/nouns/verbs + complement clause’, all structures which can be considered as metadiscursive based on the description provided above are included in the analysis. This includes oft-studied constructions like ‘VERB + *que* + complement clause’, but also other constructions such as ‘ADJ + *de* + complement clause’, ‘ADJ + *à* + complement clause’, ‘VERB + infinitive’, ‘VERB + subordinate interrogative clause’ (e.g. (10)), and ‘VERB included in a *comme*-clause’ (e.g. (11)).

- (10) *(...) il a juste été possible de **constater** si le locuteur en place luttait afin de conserver sa place de locuteur principal ou non.* (FAR, 0050)
- (11) *L'avancée aurait pu être cruciale pour la question de l'endophasie, comme l'a **démontré** Vygotski, (...)* (KIAP-FR, 0020)

Cases of polysemy were also accounted for during the disambiguation process by coding the polysemous marker according to its meaning in context. The adjective *évident*, for instance, is usually used as a booster in the corpora analyzed (e.g. (12)), but sometimes takes on the same meaning as *easy* (typically when preceded by a negation, see (13)). In these cases, it was coded as an attitudinal marker.

- (12) *Premièrement, il est **évident** que des biais sont inhérents à une approche essentiellement qualitative (...)* (FAR, 0018)

- (13) *Savoir quel comportement adopter au sein d'une conversation n'est pas aisé, et les juges ont pu le remarquer. Il n'est en effet pas **évident** de savoir caractériser le comportement socialement acceptable au sein d'une conversation.* (FAR, 0098)

4.3 Statistical methods

Overall comparisons of stance marker frequencies (e.g. number of boosters) were based on the number of occurrences per text normalized per 1,000 words. As the data was not normally distributed, I opted for the Mann-Whitney U test. The test was computed in R with the function 'wilcox.test()', and the significance threshold was set at .05. Following Field et al. (2012), the Pearson's correlation coefficient r is reported as the effect size.

The number of occurrences of specific markers (e.g. *intéressant*) was too small to be properly handled by the Mann-Whitney test. Instead, comparisons of specific marker frequencies across the two corpora were based on the chi-square test of independence. The function 'chisq.test()' was used to compute the test in R (significance threshold = .05) and I followed Gries (2013) and took ϕ as the effect size. Cumulative percentages of the top 10 markers for each semantic category across the different corpora were also compared by means of the chi-square test of independence.

5. Results

This section presents the results of the analysis in terms of stance marker frequencies (Section 5.1), lexical preferences (Section 5.2) and lexical diversity (Section 5.3).

5.1 Stance marker frequencies

Table 6 gives an overview of the medians, interquartile ranges (IQRs) and mean ranks of booster, hedge, and attitudinal marker frequencies in the KIAP-FR and in the FAR. Figures 3, 4, and 5 present the boxplots and the results from the Mann-Whitney U tests for the three semantic categories.

As shown in Table 6 and Figure 2, French novices exhibit similar booster frequencies as French experts, with a median of 2.25 occurrences per 1,000 words in the KIAP-FR and 2.7 in the FAR. By contrast, French novices present a moderate overuse¹⁴ of hedges and attitudinal markers compared to French experts (see Figures 3 and 4). For example, we note a median of 1.2 attitudinal markers per 1,000 words in the KIAP-FR, compared to 1.95 in the FAR (see Table 6).

Table 6. Descriptive statistics of stance markers in French expert vs. French novice academic writing

	FRENCH EXPERT WRITING (KIAP-FR)	FRENCH NOVICE WRITING (FAR)
<i>BOOSTERS</i>	Median: 2.250 IQR: 1.575 Mean rank: 66.54	Median: 2.70 IQR: 1.3 Mean rank: 56.19
<i>HEDGES</i>	Median: 3.350 IQR: 2.45 Mean rank: 70.3	Median: 3.900 IQR: 2.5 Mean rank: 53.5
<i>ATTITUDINAL MARKERS</i>	Median: 1.200 IQR: 1.225 Mean rank: 75.12	Median: 1.950 IQR: 1.575 Mean rank: 50.01

¹⁴ In this study, the use of the terms ‘overuse(d)’, ‘underuse(d)’ and ‘significant(ly)’ means that the p-value is significant (i.e. inferior or equal to .05).

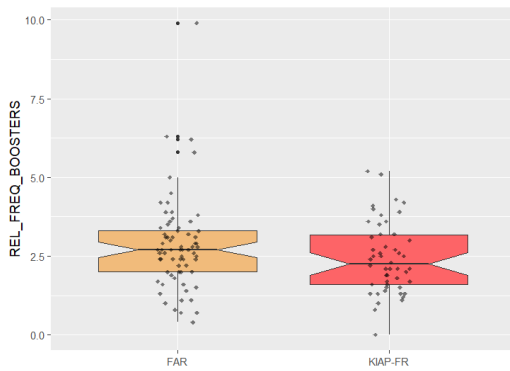


Figure 2. Boosters in French novice vs. French expert writing ($W = 1442$, $p\text{-value} > .05$; effect size: 0.16 (small))

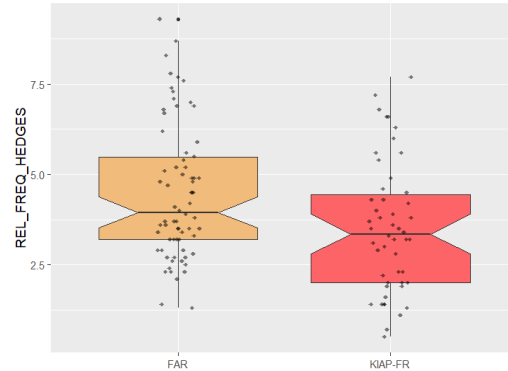


Figure 3. Hedges in French novice vs. French expert writing ($W = 1256.5$, $p\text{-value} < .05$; effect size: 0.26 (small-moderate))

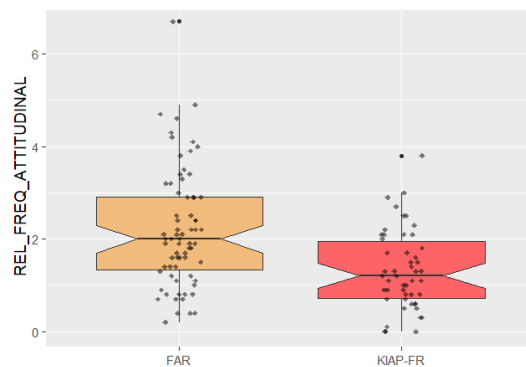


Figure 4. Attitudinal markers in French novice vs. French expert writing ($W = 1018.5$, $p\text{-value} < .05$; effect size: 0.39 (moderate))

5.2 Lexical preferences

Tables 7 and 8 present the top (i.e. most frequent) 10 boosters and hedges in French expert vs. French native novice writing. As shown in the tables, French novices' top 10 epistemic markers are quite

similar to the ones of expert writers (and mainly consist of items that are frequently used in French academic writing). First, the FAR shares eight boosters and seven hedges with French experts, such as *montrer*, *constater* and *évidemment* (for boosters) and *sembler*, *sans doute* and *hypothèse* (for hedges). In addition, both experts and novices favor verbs when hedging or boosting their claims. In the top 10 boosters, we count six and seven verbs in the KIAP-FR and in the FAR respectively, and in the top 10 hedges we note six verbs in both corpora.

The strong presence of markers like *voir* and *savoir* both in the KIAP-FR and in the FAR indicates that novices are aware of the fact that the most common markers used to express boosting in French academic writing are not the ones which convey this function in the most ‘extreme’ manner, such as *évidemment* or *bien sûr*. Instead, academic writing tends to display items that lean towards the more neutral end of the boosting continuum.

Table 7. Top 10 boosters and their relative frequencies (per 100,000 words) in French expert vs. French novice writing

FRENCH EXPERT WRITING (KIAP-FR)		FRENCH NOVICE WRITING (FAR)	
BOOSTER	REL. FREQ.	BOOSTER	REL. FREQ.
montrer	56.7	voir	42.1
voir	29.7	montrer	36.2
savoir	20.8	constater	29.2
constater	16.8	savoir	20.8
nécessairement	13.2	affirmer	19.7
évidemment	11.2	nécessaire	13.2
conclure	8.2	démontrer	11.3
affirmer	7.3	évidemment	7.6
bien sûr	7.3	conclure	7.3
nécessaire	4.6	évident	7.0

Table 8. Top 10 hedges and their relative frequencies (per 100,000 words) in French expert vs. French native novice writing

FRENCH EXPERT WRITING (KIAP-FR)		FRENCH NOVICE WRITING (FAR)	
HEDGE	REL. FREQ.	HEDGE	REL. FREQ.
sembler	62.6	sembler	120.2
considérer	61.0	considérer	80.2
paraître	26.4	penser	25.4
susceptible	21.8	hypothèse	21.1
sans doute	21.1	paraître	17.0
hypothèse	19.1	peut-être	15.1
proposer	17.1	apparaître	14.6
penser	16.1	sans doute	14.3
apparaître	13.5	estimer	12.7
certes	10.5	probablement	10.3

Despite similar overall frequencies of boosters in the KIAP-FR and in the FAR (see Section 5.1), a number of specific markers are overused by novices. It is the case of *voir*, *affirmer*, *constater*, *nécessaire* and *démontrer*. The overall overuse of hedges in French novice writing seems to be (at least partly) linked to an overuse of the markers *sembler* and *considérer*, which occupy a prominent role as hedges in the FAR (see Table 8). *Ssembler*, for instance, is about twice as frequent in the FAR as in the KIAP-FR.

With respect to attitudinal stance, the markers *difficulté*, *important* and *intéressant* are overused by French native novices (see Table 9). An analysis of the occurrences of *important* and *intéressant* in context suggests that the overuse of those markers is related to the strong use of the impersonal structure ‘il + BE + ADJ’, followed by a communication verb (e.g. (14) and (15)).

- (14) *Il est **important** de souligner que le processus de transcription a différé selon qu'il s'agissait des émissions belges ou françaises.* (FAR, 0002)
- (15) *Il est **intéressant** de remarquer que les juges ont plus facilement avancé des conditions pour qu'un comportement soit jugé acceptable (...)* (FAR, 0098)

Table 9. Top 10 attitudinal markers and their relative frequencies (per 100,000 words) in French expert vs. French novice writing

FRENCH EXPERT WRITING (KIAP-FR)		FRENCH NOVICE WRITING (FAR)	
ATTITUDINAL MARKER	REL. FREQ.	ATTITUDINAL MARKER	REL. FREQ.
difficile	14.5	intéressant	29.2
admettre	12.2	difficile	18.9
intéressant	4.9	juger	14.0
se contenter	4.6	important	13.8
juger	4.6	difficulté	9.7
prétendre	4.6	pertinent	8.4
reconnaître	4.6	préférer	6.5
difficulté	3.6	facile	4.6
problème	3.6	intérêt	4.6
particularité	3.0	essentiel	4.1

5.3 Lexical diversity

Figures 5, 6 and 7 illustrate the cumulative percentages of the top 10 boosters, hedges, and attitudinal markers respectively. For example, Figure 5 shows that, when considered together, the two most frequent boosters in the KIAP-FR (namely *montrer* and *voir*, cf. Table 7) account for about 35% of all boosters this corpus.

The 10 most frequent boosters present similar cumulative percentages in French expert writing and in French novice writing (see Figure 5; $X^2 = 1.4739$, $df = 1$, $p > .05$, $\phi = 0.03$). When added up, those 10 markers make up for 72.5% of all boosters in the KIAP-FR and 69.7% of all boosters in the FAR. This indicates similar lexical diversity levels in French expert writing in terms of booster usage. A similar finding is observed for hedges, as we note no significant differences when comparing the cumulated percentages of the 10 most frequent hedges in the KIAP-FR and in the FAR ($X^2 = 1.0475$, $df = 1$, $p < .05$, $\phi = 0.02$).

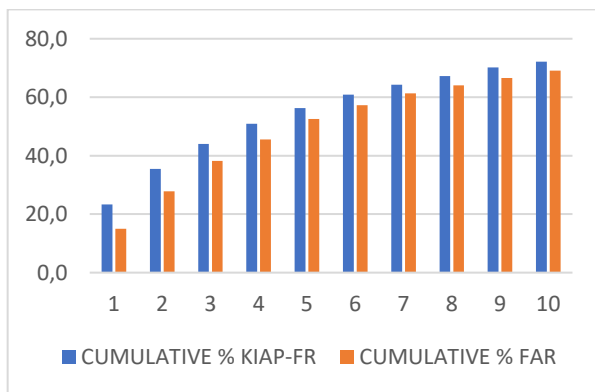


Figure 5. Cumulative percentages of the top 10 boosters in French expert vs. French novice writing

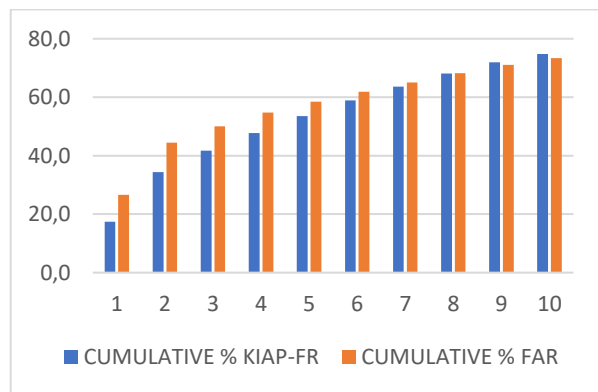


Figure 6. Cumulative percentages of the top 10 hedges in French expert vs. French novice writing

When it comes to attitudinal markers, a different tendency is observed, with higher percentages in the FAR compared to the KIAP-FR (see Figure 7; $X^2 = 19.552$, $df = 1$, $p < .01$, $\phi = 0.1$). This suggests that, when resorting to this semantic category, French novices tend to rely repeatedly on the same set of markers. This tendency is not very marked however (the top 10 attitudinal markers account for 49.1% of all attitudinal markers in the FAR compared to 45.8% in the KIAP-FR), leading to the conclusion that French

novices present slightly lower lexical diversity of attitudinal markers than French experts.

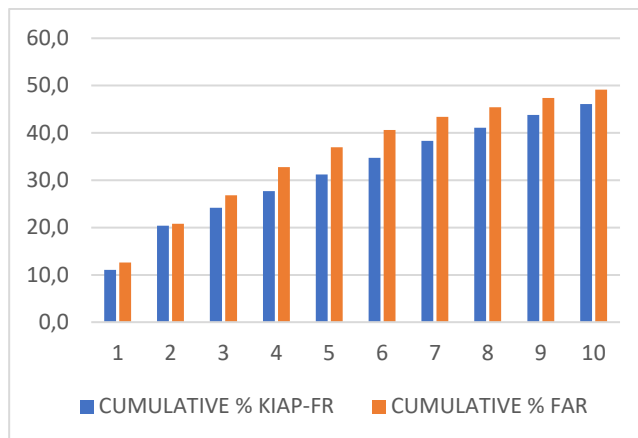


Figure 7. Cumulative percentages of the top 10 attitudinal markers in French expert vs. French novice writing

6. Discussion

While evaluative meanings have been well documented in studies on French novice academic writing, through theoretical concepts like ‘posture’, ‘ethos’, or ‘image du scripteur’ (‘writer’s image’), they have typically been examined by means of qualitative analyses of a limited number of texts, making it difficult to generalize findings. The present study showed that an approach involving larger-scale quantitative analyses of representative corpora has the potential of providing new and valuable insights into the features typically associated with French novice writing.

Results showed no significant difference in terms of booster frequencies across French expert and French novice academic writing. The study also revealed a general overuse of hedges in French novice writing, largely attributed to novices’ strong reliance

on the markers *sembler* and *considérer*. These findings indicate that French novices from the FAR adopt a more tentative tone than their expert counterparts. A sentence like (16), in which the tone is quite cautious, thus appears to characterize the texts from the FAR. This is not in line with studies from the French tradition, which have highlighted an assertive and generalizing tone from the part of novices (e.g. Delarue-Breton 2014, Rinck 2011). Such difference points to the importance of text type. The texts typically analyzed when investigating French novice writing are internship reports written by future teachers. This text type might encourage students to take an authoritative stance, as they are supposed to comment on the observations they made in class. Such circumstances “are likely to reinforce the tensions between (...) the competing notions of objectivity and subjectivity” (Colin & Dolignier 2019: 40). French academic papers as analyzed in the present study, by contrast, are engrained in a culture of ‘effacement énonciatif’, in which writers are expected to remove themselves from their texts – at least to a certain extent (e.g. Tutin 2010). French novices apparently understand the importance of this requirement, and perhaps apply it to an extreme. Another possible explanation for French novices’ tentative tone is linked to the fact that student writers occupy an intermediate position in the academic sphere, “at the intersection between two socio-institutional areas”, namely academic education and research (Reuter 2004: 10). For example, students’ written productions are subject to a reader-evaluator, who assesses the quality of their work. From this position can emerge a feeling of ‘writing insecurity’ among student writers, who might not be as confident as their expert counterparts in advancing hypotheses, presenting results, etc. (Colin & Dolinger 2017). These circumstances could thus lead to a stronger reliance on hedging devices compared to expert writers.

- (16) *Du point de vue du « paraître », l'homme **semble** se rapprocher d'une attitude **considérée** comme typiquement féminine : celle du soin du corps.* (FAR, 0026)

The writers from the FAR corpus were shown to adopt a more personal style than French experts, through an overuse of attitudinal markers such as *important* or *difficulté*. Qualitative studies of French novices' texts have already observed such an attitude-laden tone (e.g. Rinck & Pouvreau 2009, Reuter 2004). The findings from the present study therefore provide quantitative evidence to generalize this feature. This personal tone is combined with more wordiness from the part of novices, who make frequent use of expressions like *il est important de souligner* or *il est intéressant de noter*. For example, a sentence like (17) could easily be expressed in a more concise, denser manner, by simply removing the sequence *il est important de souligner que*.

- (17) *En outre, il est **important** de souligner que, parmi ces six interviews, trois proviennent de la radio publique belge « La Première » et les trois autres sont toutes issues de « France Inter », radio publique française.* (FAR, 0002)

This study also highlighted additional stance features in French novice writing. We noted slightly lower attitudinal marker lexical diversity in French novice writing compared to French expert writing. French novices were also shown to rely heavily on specific markers, like *voir*, *sembler* and *considérer*. This indicates that the French novices in the corpus under study rely on a more limited lexical repertoire of stance markers compared to French expert writers, and that their written productions are characterized by the use of lexical 'teddy bears' (Hasselgren 1994, Nesselhauf 2005). Despite this limited repertoire, and as revealed by the analyses of lexical preferences, French novices appear to exhibit general good command of common epistemic markers. This is shown by their

strong reliance on epistemic verbs and by the high number of shared epistemic items identified between the KIAP-FR and the FAR (eight boosters and seven hedges out of the 10 most frequent ones). These findings might indicate a high level of academic writing acculturation from the part of FAR writers.

7. Conclusion

The aim of this study was to apply a large-scale quantitative approach to stance to corpora of French academic writing, thereby revisiting some of the findings commonly highlighted in the framework of smaller-scale qualitative research. The results help to nuance features typically associated with French novice writing. For instance, while the adoption of an assertive tone by French novices was not confirmed by the results, the present study did confirm the adoption of a personal tone from the part of those writers, making it possible to further generalize this finding.

Such approach also highlighted findings that are not typically emphasized in studies on French novice writing. Analyzing the data quantitatively has made it possible to find, for instance, slightly lower lexical diversity of attitudinal markers in French novice writing compared to French expert writing. A good command of epistemic lexical preferences in novice texts was also observed, as well as more wordiness.

8. References

- Aull, L. (2015). *First-Year University Writing: A Corpus-Based Study with Implications for Pedagogy*. Houndmills, Basingstoke, Hampshire ; New York: Palgrave Macmillan.

- Biber, D. (2006). *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*. John Benjamins, 201-226.
- Biber, D., Johansson, S., Leech, G. Conrad, S. & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. Longman.
- Blanche-Benveniste, C. (2002). Phrase et construction verbale. *VERBUM*, XXIV(1), 7-22.
- Colin, D. & Dolignier, C. (2017). L'auctorisation d'étudiants de lettres dans l'écriture d'un mémoire de master enseignement. *SCRIPTA, Belo Horizonte*, 21(43), 208-233.
- Colin, D. & Dolignier, C. (2019). Le mémoire d'initiation à la recherche en formation d'enseignants de lettres : Entre posture de praticien et positionnement scientifique. *Le français aujourd'hui*, 204, 37-50.
- Delarue-Breton, C. (2014). Créativité, au(c)torisation et dialogisme : le mémoire de master MEEF, miroir de l'expérience psychique ? *Congrès Mondial de Linguistique Française*.
- Deronne, C. (2011). Postures énonciatives du scripteur : le cas du mémoire d'IUFM. *Pratiques*, 149/150, 91-111.
- Donahue, C. (2002). Effets de l'écrit sur la construction du Sujet textuel à l'université. *Spirale*, 29, 75-108.

- Field, A., Miles, J. & Field, Z. (2012). *Discovering Statistics Using R*. SAGE.
- Fløttum, K., Dahl, T. & Kinn, T. (2006). *Academic Voices*. John Benjamins.
- Fløttum, K. & Vold, E. (2010). L'éthos auto-attribué d'auteurs-doctorants dans le discours scientifique. *Lidil*, 41, 41-58.
- Granger, S. & Paquot, M. (2010). The Louvain EAP Dictionary (LEAD). In A. Dykstra & T. Schoonheim (eds.), *Proceedings of the XIV EURALEX International Congress*. Leeuwarden, The Netherlands, 321-326.
- Gray, B. & Biber, D. (2014). Stance markers. In K. Ajimer & C. Rühlemann (eds), *Corpus Pragmatics: A Handbook*. Cambridge University Press, 219-248.
- Gries, S. Th. (2013). *Statistics for Linguistics with R. A Practical Introduction* (2nd edition). De Gruyter Mouton.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237-259.
- Hyland, K. (2005a). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.

- 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 et al. (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.
- Kilgariff, 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.
- Molinier, C. & Levrier, F. (2000). *Grammaire des adverbes. Description des formes en –ment*. Droz.
- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Pho, P.D. (2013). *Authorial Stance in Research Articles: Examples from Applied Linguistics and Educational Technology*. Houndmills, Basingstoke, Hampshire ; New York: Palgarve Macmillan.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Rinck, F. (2011). Former à (et par) l'écrit de recherche. Quels enjeux, quelles exigences ? *Le français aujourd'hui*, 174(3), 79-89.

- Rinck, F. & Pouvreau, L. (2009). La mise en scène de soi dans un écrit d'initiation à la recherche en didactique du français. *SCRIPTA*, 13(24), 157-172.
- Reuter, Y. (2004). Analyser les problèmes de l'écriture de recherche en formation. *Pratiques : linguistique, littérature, didactique*, 121/122, 9-27.
- Tutin, A. (2010). *Dans cet article, nous souhaitons montrer que...* Lexique verbal et positionnement de l'auteur dans les articles en sciences humaines. *Lidil*, 41, 15-40.
- 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*. Université de Bergen.