Comparing indicators of authorial stance in psychology students' writing and published research articles

Alice Henderson, Robert Barr

Henderson, A . et R. Barr (2010), « Comparing indicators of authorial stance in psychology students' writing and published research articles », *Journal of Writing Research 2*(2), 245-264.

Comparing indicators of authorial stance in psychology students' writing and published research articles

Alice Henderson & Robert Barr

University of Savoie | France

Abstract: This article presents the results of a pilot study examining the use of first-person pronouns, certain adjectives and grading adverbs in a corpus of 51 French psychology student papers written in English as a second language. These results were compared to a corpus of published psychology articles and to a sub-corpus of psychology student texts from the British Academic Written English corpus (BAWE). Strategic use of pairs of evaluative words was found in the students' texts but not in the published texts. However, the variables of native language and level of field expertise cannot explain all of the variance observed. Future work will improve the validity of the findings by using larger corpora of student and published texts.

Keywords: authorial stance, psychology, English, academic writing

Written academic English obeys rules concerning the use of formulaic expressions (Jones & Haywood, 2004). Such series of words play diverse roles within a text and in the larger discourse community (Swales, 1990). Because members of these communities respect the rules (Dressen-Hammouda, 2008), writing can reflect (at least partially) the degree of a writer's discourse community membership.

Authorial stance is the author's viewpoint on the material to which they are referring. Biber defines it as "the ways in which an author or speaker overtly expresses attitudes, feelings, judgments, or commitment concerning the message" (1988, p.204). Evidence of stance is found in lexical items (e.g. adjectives expressing evaluation, or choice of reporting verb), which are easily found in a corpus using simple text analysis software. More importantly for the aspiring writer in a foreign language, many of these are embedded in formulaic sequences. This term is hard to define, but in *Formulaic Sequences: Acquisition, Processing and Use,* Schmitt & Carter (2004) opt for Wray's definition:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (2002, p.9, cited in Schmitt & Carter, 2004, p.3)

Other researchers refer to bundles and clusters, linking their use to both language and disciplinary expertise. Hyland uses the un-connoted term "bundle" to refer to "words which follow each other more frequently than expected by chance, helping to shape text meanings and contributing to our sense of distinctiveness in a register. ... the absence of such clusters might reveal the lack of fluency of a novice or newcomer to that community" (2008, p.5)". Cortes agrees that "The frequent use of lexical bundles, for example, seems to signal competent language use within a register to the point that learning conventions of register use may in part consist of learning how to use certain fixed phrases" (2004, p.398).

Students writing in a foreign language have a notoriously difficult time acquiring a native-like feel for recurrent expressions. They may over- and/or under- use such expressions (Meunier & Granger, 2008: Nesselhauf, 2005). This writing problem may stem from difficulties learners have reading in English as a foreign language. Such learners may be unaware of authorial stance and the norms of disciplinary culture because when they read they may be painstakingly deciphering at word level, often reading word by word, from start to finish. Even native speaking (NS) English students often fail to acquire and use the variety of lexical bundles appropriate to published academic writing, despite extensive reading (Cortes, 2004, p.413). Therefore, students in general need guidance in noticing how and where stance appears in texts. Once students have noticed the expressions authors use to signal their presence, to interact with other authors and to express membership of their community, it should be easier for students to grasp that such

interaction exists. This is arguably one of the first steps in understanding the exchange of ideas upon which critical reading depends and which is the basis of academic writing.

The assumptions underlying the present study are that if students can become more aware of some of the language used to express authorial stance in the texts they read, they will achieve two things: improve their mastery of that language in their own writing and take another step toward becoming members of a discourse community (as defined by Swales, 1990). Thus, analysis of their writing may provide useful insights into the novice-expert continuum of membership in a discourse community.

This study fills a gap in existing research because it looks at writing in the field of psychology; an on-line search of *Journal of English for Academic Purposes* and *English for Specific Purposes* from 1988 onwards shows that writing in many other fields has been studied (e.g. history, literary criticism, sociology, medicine, biology, pharmacology, engineering, politics, materials science, agriculture, applied linguistics, biochemistry, philosophy, management, organic chemistry, computer science, wildlife behavior, conservation biology, law, nursing.). In addition, although sections other than the introduction of the research article have received a fair amount of attention, no published work was found that focused on the introduction of psychology research articles and/or on psychology students writing introductions in English as a non-native language.

1. Method

1.1 Teaching task

The teaching task (see Appendix A undoubtedly had an effect on the writing the French university students produced in English. The task began with noticing exercises in which the students were required to identify statements of authorial stance in a short French text. Then they had to read the introductions of psychology articles published in English and find examples of expressions where authors explain how their contribution builds on or breaks from existing research (Flottum, Dahl, & Kinn, 2006; Boch, Grossmann, & Rinck, 2009). They then read published articles in English and created their own lists of further evaluative terms (adjectives, adverbs, and bundles). The French students were explicitly instructed to use these words in the subsequent writing task: writing a literature review or introduction to a research article. The instructions and the amount of time devoted to explicit work with these expressions were intended to encourage students to use the expressions effectively. A teaching effect was therefore expected in the results.

1.2 Corpora

Three collections of English text were compared, as shown in Table 1: non-native speaker (NNS) French psychology students' texts, psychology research article introductions from the DOAJ (Directory of Open Access Journals) and native-speaker (NS) psychology students' texts from the BAWE corpus. The BAWE student writers were in their 3rd and 4th year at university. The French students were in their final year of a three-year university Psychology course and their texts were used only with their permission. All texts were written by individuals but seven of the BAWE students provided more than one text. The DOAJ texts were selected as being representative of the types of research article introductions French psychology students read in English.

Table 1. Text corpora used in the study

	NNS French psychology student texts	Psychology article introductions from DOAJ	NS psychology student texts from BAWE corpus
Number of writers/texts	51/51	15/15	7/18
Mean length of texts, n° of words	870	897	2334
Total n° of words	46,084	12,837	41,454

Table 2 provides more detail about these texts. The texts come from a variety of psychological fields that adhere to different epistemological and methodological traditions and use more than just the IMRAD text structure (Introduction-Methods-Results-Analysis-Discussion). Nonetheless, in the texts selected here from

a variety of sub-disciplines, the introduction section arguably fulfils the same rhetorical and discursive functions, namely to show that the author has understood the existing research and is able to situate their own research in this context. This section of a research article is therefore a potentially fruitful place to look for evidence of authorial stance.

Table 2. Corpus of published academic writing in Psychology

DOAJ Journals & n° of texts from each	Australian Journal of Educational & Developmental Psychology (4) Psicologia: Teoria e Pesquisa (5) Anales de psicología (3) @ctivities (3)
Years	2000-2003, 2006, 2007
Type of texts	Introduction sections of research articles
n° of texts/writers	15/15
Mean length of texts	897 words
Total n° of words	12,837

One uncontrolled variable is the native language of authors. The BAWE texts were all written by native English-speakers and the NNS texts by native French-speakers, but it was not possible to determine the native language of the DOAJ authors. However, it was assumed that journal reviewers accept writing that conforms to an implicitly defined native-speaker norm.

In general, the design of corpora must take into consideration issues of size, content, representativeness and permanence (Hunston, 2002). In terms of size and content, the current study tried to compare like with like, as is shown in Table 1. Quantitatively, these corpora are quite small and hence far from representative; therefore none of the linguistic analyses are statistically valid. Nonetheless, as this is merely a pilot study, occurrences were standardized for 10,000 words to make it easier to compare results. A major qualitative difference relates to genre, as the BAWE sub-corpus does not include extended introductions or literature reviews as separate genres, but rather texts from several “genre families”: critique (3), essay (12), explanation (1), proposal (1), and empathy writing (1). While it would be difficult to foresee how and to what extent this genre mismatch affects results, it is taken into account in their interpretation. The selected genre families from the BAWE exhibit functions found in the literature review, such as comparison and evaluation. Therefore, BAWE texts have been included in order to enrich comparisons of the mastery of written English and level of field expertise, as evidenced in specific lexical features.

1.3 Corpus processing and analyzing procedures

The French students’ writing was transformed into .xml files using the <oXygen/>XML editor as part of the protocol for the larger Scientext corpus project. The DOAJ psychology articles were originally .pdf files and were transformed into .txt files. All files were manually checked or “cleaned” for spelling errors, omissions, etc. The BAWE student writing was extracted from the larger BAWE corpus. These were also transformed into .txt files and checked. None of the texts were annotated, for example, for syntactic features. The corpora were analyzed with *AntConc* (Anthony, 2007), a freeware corpus analysis toolkit which works best with texts in .txt or .xml format. *AntConc* generates concordances (a list of the occurrences of a word including its context), clusters, collocates, keyword lists and frequency lists. For this study, *AntConc’s* term “N-gram” is used synonymously with Hyland’s term “bundle”. First, a list of N-grams was generated to give an idea of the frequency and variety of formulaic sequences or bundles. Concordances of these N-grams were then used to hand-sort N-grams by function, in relation to whether or not they express authorial stance. In a second step, concordance searches were used to refine a predetermined list of search words. As this is merely an exploratory study, a very short list of gradable adjectives and adverbs was drawn up: *old, new, high, low, rather, fairly, more, most, so, too, very*. This list is based on words referred to in other studies (Hyland, 2002: Hunston & Sinclair, 1999: Nesselhauf, 2005: Meunier & Granger, 2008) and on teachers’ experience of the simplest adjectives and adverbs commonly used by French students writing in English. Such a short list is easy to search for with *AntConc.* Nevertheless:

> An adjective which has comparative and superlative forms and which is sometimes or often used with a grading adverb ... is likely to be evaluative, though it is not necessarily so. ... gradedness indicates

comparison, and comparison with a norm or scale is often a matter of subjectivity. Subjectivity is one of the contributors to evaluative meaning. (Hunston & Sinclair in Hunston & Thompson, 1999, p.92)

Looking at the concordances for the adjective collocates of such adverbs helped to reveal which adjectives to focus on in further concordances. The adjectives *important* and *present* were also included in the adjective list in order to test teachers' perceptions that they are frequently misused by French students.

2. Results & Analysis

In this section, N-grams are presented first as an initial quantitative approach to the corpora. This is followed by the qualitative detail of concordances. Occurrences were excluded if they occurred in a quotation, in a proper noun (e.g. New York) or in a questionnaire item (e.g. *Do you typically spend time with your children in the evening or only at weekends?).*

2.1 N-gram data

Table 3 shows the data for the five most frequent three-item N-grams, where there is a noticeable lack of stance expressions. The vocabulary of "functions and means/cause and effect/result" was explicitly taught and yet only *in order to* shows up in NNS writing, though *the role of* could refer to a causal relationship. The BAWE students used *due to the* and *in order to* proportionately much more frequently than the NNS student writers. The additive *as well as* bundle appeared only in the NNS writing but is embedded in the "best" 4-item bundles which Hyland argues a general EAP course should cover (2008): *On the other hand, As well as the, In the case of, The end of the*. Perhaps the prevalence of these four bundles is due to the simplicity of their functions: establishing a comparison, adding another element, pointing to an example or pointing to the final part of something. One expression in Table 3 which identifies a text as being academic is *can be seen*, which is normally used to refer to results and what they indicate. It should be "frequent and unremarkable" (Hyland, 2008, p.5) but only the BAWE students used it. The BAWE texts also included the only occurrences of a passive (*can be seen*), which is astonishing given that the passive voice focuses attention away from the agent and this is supposed to be typical of academic text. However, it may be more typical of a results section, which neither the DOAJ nor NNS texts include. All the lists of 3-item bundles include noun phrase + post modifier fragments: *the role of, a number of, a lot of*. Two of the five 3-item bundles from the DOAJ introductions were lexically quite specific (*transition to adulthood, of the family*). Only the analysis of a much larger corpus could determine whether or not NS writers tend to use more lexically rich bundles in introductions. The bundle *as cited in* explicitly refers to source attribution. It only appeared in the BAWE texts even though it would be expected to occur frequently in research article introductions. Its absence indicates either that other means of source attribution were used or that sources were not referred to. The word *according* should also be an obvious means of referring to a source. However, concordance data reveals that *according* was used for this purpose 5 times in the BAWE corpus and 7 times in the DOAJ corpus, which has almost three times fewer words. One of the French students used *according to me* (1 occurrence), showing that they accepted their teacher's advice to avoid the expression, though they may or may not accept that it explicitly confers too much authority to the validity of the writer's ideas. Source attribution was therefore being expressed via other means.

Table 3. Five most frequent three-item N-grams: raw n° of occurrences

NNS texts	DOAJ texts	BAWE texts
a lot of (19)	transition to adulthood (15)	as cited in (40)
in order to (15)	of the family (13)	in order to (29)
as well as (12)	the transition to (13)	can be seen (27)
the role of (13)	a number of (9)	due to the (25)
there is a (13)	in order to (7)	there is a (21)

Table 4 shows data for the five most frequent four-item N-grams. The most puzzling absence is that of *on the other hand*, which is Hyland's most frequent 4-item bundle in biology, applied linguistics, electrical engineering and business studies (2008). Given the fact that simple contrasts are a common structure in literature reviews - or arguably many texts where different points of view are compared - this absence

could be attributed to the relatively small size of the corpora, compared to Hyland's 3.5 million word corpus.

Table 4. Five most frequent four-item N-grams: raw n° of occurrences

NNS texts	DOAJ texts	BAWE texts
at the time of (9) We can say that (9) the child's self-perception (8) been found to be (7)	x	can be seen that (11) it can be seen (10) et al found that (8) knowledge and phonological awareness (8) that there is a (8)

2.2 Pronouns

Concordance data is key when trying to clearly identify authorial stance, especially in relation to pronouns. Raw frequency data is insufficient to identify stance, partly because writers often use pronouns when they are merely highlighting the organizational structure of their text. This function dominates both the NS and NNS student texts, as they tend to clearly outline what they are going to do or what they have done, or justify the choices of texts/subjects. Therefore the author is quite obvious in these texts, but more as a "signposter" and less as an "expert" who analyses and evaluates ideas/schools of thought, etc. This is a "fairly low risk writer role", according to Hyland (2002, p.1100). Table 5 shows the variety of pronouns used by the different writers. The most surprising finding concerns the prevalence of first-person pronouns. Textbooks do exist which encourage writers to use the first person in order to make their personal voice clear (Hyland, 2002) and the French students may have received conflicting advice from different teachers. However, other explanations may be found by looking at the five functional categories into which the occurrences in the current study fall:

1. Signaling text structure, such as *First, I will stress that..., In a second I will develop ...* This occurred 12 times in the NNS texts, 10 times in the BAWE texts and 7 times in the DOAJ introductions.
2. Justifying methodology, such as *I have selected three studies because....*: This occurred 5 times in the NNS texts, but never in the BAWE texts or in the DOAJ introductions.
3. Explicitly claiming expert or "non-layman" status, such as these examples from the NNS texts: *Being psychological student, I wanted to know what is the real* and *Hence, for my future profession, I take the opportunity to exploit this subject.* This occurred 3 times in the NNS texts, but never in the BAWE or DOAJ texts; it would be interesting to use a larger corpus to see if "real" experts ever explicitly define themselves as such.
4. Expressing what has been understood, "showing" knowledge, for example *I believe that PTSD symptoms are associated with* and *Indeed, I guess that self-esteem (SE).* This is also where the most subjective verbs were found (*I feel, I believed, I asked myself*), verbs that are rarely used in academic texts. There were 6 occurrences in the NNS texts. The BAWE texts include one use of *I feel* and 9 uses of *I believe*. However, all 9 occurrences of *I believe* came from two student writers: one student produced 4 occurrences and the other produced 5 occurrences. This latter student also provided the sole occurrence of *I feel*. Consequently, any conclusions drawn from such results must take into account the small size of the BAWE corpus.
5. Listing events (*I selected, I have chosen, I was able to)*: This occurred 9 times in the NNS texts and 3 times in one of the DOAJ texts, where the author explains how he came to be involved in this research.

Table 5. Pronouns: raw n° of occurrences and n° of occurrences per 10,000 words

Item	NNS texts (46084 words)		DOAJ texts (12,837words)		BAWE NS texts (41454 words)	
	Raw	per 10,000 words	Raw	per 10,000 words	Raw	per 10,000 words
I	35	7.59	10	7.79	20	4.82
me	7	1.50	0	0.00	1	0.24
my	15	3.25	3	2.33	3	0.72
we	153	33.20	30	23.37	63	14.20

our	31	6.73	17	13.24	43	10.37
us	24	5.21	6	4.67	15	3.62
you	4	0.86	0	0.00	5	1.21
your	3	0.65	0	0.00	0	0.00

The four-item N-grams in Table 4 yield the first examples of the first person pronoun *we*, used only by the NNS students. Concordances for the pronoun *we* occur in several expressions: *we observe, we could conclude saying that, we will see primarily, we can read, we can find, we can deduce, as we have noticed, the differences that we observe, we can also observe*. These seem to be translations of pronouns which are commonly used in French academic writing, even by single authors: *on* (1st person) and *nous* (3rd person), which is used here as a "pluriel de modestie" as opposed to the "nous de majesté". Another possible explanation is that the NNS students were trying to avoid using "I", which they have been told by their English teachers clashes with the supposedly neutral, objective, replicable nature of academic & scientific texts. The concordance data on pronouns seems to support the likelihood of a combination of factors. Similarly, in Table 5 the large number of occurrences of *we* in all the texts may reveal transfer from the NNS' native French. However, the fact that it is used by both the native and non-native writers and by both field experts and novices would support the idea that conventions concerning first-person pronoun use are not stable (Hyland, 2002, 1095). The high numbers of occurrences for *us* and *our* in the NNS texts where the writer is directly addressing the reader could be due either to transfer or to this instability: *What could be said beyond all that? What do you think when the hypothesis...* The other three NNS uses of *you*, which came from three different student writers, express no stance because they could be replaced by a passive construction or *one*: *you have differences ways to study something*, *you can or not be vulnerable to get this illness, It's interesting when you know that cardinal symptoms.*

In the NNS texts the pronoun *we* was used with a modal verb in 74 of the 153 occurrences. This is important because modals may express stance. With the exception of two direct questions to the reader (*How can we explain this phenomenon?, Should we establish a parallel between animal*) these *we* + modal examples can be categorized into five categories:

1. Signaling text organization: 19 occurrences (*to begin we will define, now we will have an interest for, Third we shall see*), and 12 occurrences solely concerning conclusion (*In conclusion we can see that, To conclude we can notice*). This category also includes 5 comments such as *To answer this question we will see, we will focus on, we will present.*
2. Comparing: 11 occurrences (*but, on the contrary, like this, on the one hand*)
3. Referring to other research: 2 occurrences (*among these articles, with reference to these studies)*
4. Expressing causal relations (*therefore, thus, that's why, so*): 5 occurrences
5. 20 direct translations from the French (*nous pouvons dire, nous pouvons observer)* which could easily be removed or replaced by passive constructions, without weakening the idea (*we can think/say/find/observe*).

In the BAWE texts these categories do not cover all 21 occurrences of *we* + modal, such as the 5 *occurrences* where *we* is used to make recommendations (*we should work towards, we could greatly reduce*). However, there were no occurrences for the text organization and comparing categories. Nine occurrences of *we* in the BAWE texts could be replaced by passive constructions (*cannot do so/may simply/cannot see into the future)* but would lose much of their rhetorical impact: *the one question we have to answer is 'where does the 'me' stop?'* has more impact than the passive constructive *the question which has to be answered*. The DOAJ introductions contained only 6 examples of *we* + modals, one of which makes a recommendation (*we should focus*), three suggest definitions (*we could speak of, we could define, we can define*), and two simply list actions (*we can infer, we have to emphasize*) that could arguably be replaced by passive constructions.

In summary, in all three corpora most occurrences of *we* were simply listing events (*we analyzed, we chose, we found*) without any evaluative connotations. The BAWE students used *we* less frequently than the NNS students, and their uses of *we* +modal were more central to the progression of ideas in the text and less about explicitly signaling text structure.

2.3 Grading Adverbs

Grading adverbs are used with adjectives to show that something or someone has more or less of a quality, for example: *childhood obesity is a very serious and rather urgent issue*. The variety of grading

adverbs used by the different writers is shown in Table 6. The high number of overall occurrences for *more* and *most* can be explained by their use in describing data. The high number for *so* in the NNS texts can be attributed to the 26 occurrences where it is used as a synonym for *therefore*. Twenty of the 25 overall occurrences of *rather* in the BAWE texts are found in *rather than* constructions. The frequency of use of *rather* and *more* was markedly lower for the NNS student texts. However, these raw frequency figures do not reveal the feature of most interest to the present study: authorial stance. Therefore, concordances were generated for each adverb. The figures in Table 7 indicate the total number of stance-oriented occurrences compared to the total number of occurrences in the corpus.

Table 6**.** Grading adverbs: raw n° of occurrences and n° of occurrences per 10,000 words

Item	NNS texts (46084 words)		DOAJ texts (12,837words)		BAWE NS texts (41454 words)	
	Raw	per 10,000 words	Raw	per 10,000 words	Raw	per 10,000 words
rather	7	1.52	7	5.45	25	6.03
fairly	0	0.00	0	0.00	3	0.72
more	114	2.39	45	35.05	126	30.40
most	61	13.24	23	17.91	38	9.17
So	36	7.81	9	7.01	44	10.61
too	4	0.87	2	1.56	3	0.72
very	40	8.68	12	9.35	20	4.82

Table 7. Stance-oriented grading adverbs: raw n° of stance-oriented occurrences compared to total raw n° of occurrences

Item	NNS texts	DOAJ texts	BAWE NS texts
rather	1/7	2/7	3/25
fairly	0	0	2/3
more	27/114	9/45	36/126
most	23/61	17/23	19/38
So	6/36	2/9	2/44
too	¾	0/2	0/3
very	40/40	6/12	20/20

In terms of stance, the most remarkable figures are the ones for *very*, which both groups of student writers always used in a statement of stance: *Although this framework tells us very little about ..., These studies also told us very little about the ...,* and *However, there has been very limited research as to* (from the BAWE students); *eating disorders recently increased in a very scary way ..., the role of school which is not very developed by the authors...* and *psychoneuroimunology is a very important subject of research today* (from the NNS students). In the DOAJ introductions, 6 of the 12 uses of *very* expressed stance, for example: *differences appear to be very unreliable..., very few studies exist..., So far, we know very little.*

However, the categorizing of these concordance examples revealed that student writers were pairing stance indicators with other evaluative terms. Pairing was found only once in the DOAJ introductions (*for a number of years with, regrettably, very little response*). Examples of pairing from BAWE texts include:

a) *it can be seen* that psychology is *very* much part of the debate over whether

b) *Therefore, although more* empirical work is *required*, *more* theoretical work *would also be needed*

c) *Therefore it may be more useful* to use these concepts in conjunction with

d) *I personally believe* reduced-inhibition to be the *most plausible* and *substantiated* theory for age-relate

Examples of pairing from NNS texts are less idiomatic but nonetheless reveal the author's position:

e) *Thereby the most* hard in an eating disorders' therapy

f) *Nevertheless*, the *most* important thing is that prisoners'

g) possibilities *more* and *more* subtile, so *the most important* from *my point of view*

h) *We have to* note that the *more* severe pathology is,

Such "lexical pairing" may be a strategy for expressing stance using a limited vocabulary. For example, the sole DOAJ example pairs *very* with *regrettably*; however, such rare adverbs as *regrettably* are probably not as available to student writers, especially non-native writers.

In the BAWE examples a) and c) the presence of the author is hidden by the use of anticipatory-IT constructions. On the other hand, the NNS examples g) and h) and the BAWE example d) explicitly use first-person pronouns, highlighting the author's presence. The use of the modal of obligation (*we have to*) makes this presence even stronger in h), In d) the limiting adverb *personally* is ambiguous; it either tempers or reinforces the strength of the *I* pronoun because it emphasizes the separation of the author's belief from others'.

2.4 Adjectives

Occurrences of the selected adjectives used by the different writers are shown in Table 8. An extremely limited number of occurrences of the simplest adjectives indicate stance. All the examples of *old* from the NNS students refer merely to the age of study participants, with the exception of *It's important to note that their criticism is quite old and that from the investigation have progressed*. Similarly, *high* always refers to rates, levels or high school. None of the occurrences of *low* refer to authorial stance (*low levels, low self-esteem*).

Table 8. Adjectives: raw n° of occurrences and n° of occurrences per 10,000 words

Item	NNS texts (46084 words)		DOAJ texts (12,837words)		BAWE NS texts (41454 words)	
	Raw	per 10,000 words	Raw	per 10,000 words	Raw	per 10,000 words
new	12	2.60	23	17.91	23	5.55
Old	1	0.22	3	2.34	5	1.21
low	30	6.5	10	7.79	10	2.41
high	37	8.03	20	15.58	14	3.38
important	66	14.32	17	13.24	48	11.58
important*	5	1.08	0	0.00	0	0.00
present	12	2.60	11	8.57	18	4.34
present*	6	1.30	0	0.00	0	0.00

* indicates non-standard use

Overall, in terms of stance, *new* is a much more productive item for both the NNS students' texts and the published DOAJ introductions. Strictly speaking, the French students are not using *new* to express authorial stance:

i) It could be the object of a *new* research.

j) To finish and to introduce a *new* framework, it is worth emphasizing that

These uses of *new* are similar to text-oriented bundles, as they mark transitions, mark results, organize stretches of discourse or frame arguments (Hyland, 2008, p.13-14). Even though the term is being used to refer to a single adjective and not a bundle, examples i) and j) reveal how the students are trying to take their audience into account, without going so far as explicitly stating their own attitude or evaluation. They are structuring and showing that they see how studies fit together but they do not want to adopt a riskier, more evaluative tone. In contrast, in the DOAJ introductions 11 of the 23 uses of *new* occur in wider contexts which seem to be more participant-oriented, for example:

k) I will discuss the need and possibility for a *new* type of network intervention

l) has the potential to open the door for a *new* line of empirical investigation.

Thus, *new* was used less frequently by the NNS writers than by the BAWE and DOAJ writers. The student writers embedded *new* in larger, non-evaluative series of words that signaled discourse structure or described how research fit together.

Although the words *important* and *present* can indicate transfer between the student writer's native French and English, they were included in this study because they can also express authorial stance. Despite teachers' perceptions that they would be used incorrectly, only half of the uses of *present* in the NNS texts show such transfer, for example:

m) this consumption stays very *present* in the teenagers' population.

n) a phenomenon more and more *present* in our society today

o) Thus social factors are *present* and influencing in the universe of the child.

However, examples m) – o) clearly express the writer's stance concerning the potential impact of the consumption, phenomenon and belief. Concerning the word *important*, which can be used in French to express quantity as well as quality, only five of the 66 occurrences represent direct translations, for example:

p) a lot of crisis provoked by an *important* anxiety

q) but also an *important* risk of the athletes health

r) they suffer from an *important* professional embarrassment:

In French it is also possible to use adjectives as nouns, leading to: *The most important from my point of view*. Even though this example is grammatically incorrect, it clearly expresses the NNS writer's authorial stance.

Adverbs and adjectives were searched for using the collocates function of AntConc and the results are presented in Table 9. "Type" refers to the different units and "token" indicates the number of occurrences for those units. This distinction is important because NNS writers are assumed to use fewer types of adjective (*small, good, high*) but use them many times, whereas a NS writer is more likely to use a more varied vocabulary (more types) but fewer adjectives overall (tokens). Surprisingly, Table 9 shows that the NNS students used 47 types in single occurrences, compared to only 12 for both the DOAJ and BAWE texts. This is all the more striking because, although the DOAJ corpus is much smaller than the NNS and BAWE corpora, the NNS and BAWE corpora are almost the same size. Some of the NNS collocates were non-standard English (*more present, more subtile, more disseminable*) but overall, the French students did not avoid qualifying nouns and verbs. This becomes even more obvious in collocates with 2 or more occurrences. The NNS students used a respectable total of 18 types in their texts, compared to 25 types for the BAWE texts and 12 types for the DOAJ introductions. The concordances show that, as in the DOAJ, the BAWE students did not use the epistemic adverbial to express stance:

s) Additionally, ... some authors have argued that some types of natural disasters are *more likely* to evoke symptoms

t) In the same way, Shannon, Lonigan, Finch & Taylor (1994) found that children younger than 13 were *more likely* to be affected by

Table 9**.** Adverb+Adjective collocates: raw n° of types and raw n° of occurrences (tokens)

n° of occurrences (tokens)	One	Two	Three	Four	Five or more
NNS texts, n° types	47	10	5	2	1
DOAJ texts, n° types	12	12	9	2	2
BAWE NS texts, n° types	12	10	2	0	0

Examples t) simply describes what other researchers did, by referring to observable, quantifiable findings. Example s), however, does contribute to expressing stance because it is paired with *argued* instead of with a less evaluative verb such as *said* or *stated*.

The results and analysis show how using predetermined word lists, word list generators and concordance data can be complementary. First, word lists were generated to determine which adjectives

and adverbs had been used, and to verify that the items on the predetermined list did occur. Secondly, concordances were generated in order to see the context in which words appeared. Examples of stance-oriented lexical items could then be manually extracted from the concordances. Evidence of lexical pairing as a means of subtly expressing authorial stance may not have been noticed without the automatic generation of word lists.

3. Discussion

The analyses revealed formulaic sequences that were unforeseeable because they were non-standard English. This highlights the importance, when working on non-native speaker corpora, of using the corpus data in addition to previously determined search lists. Similarly, it would have been difficult to search for all the possible lexical pairings of evaluative terms. Starting with a list of grading adverbs meant that many such pairs were found; however, starting a search with modal verbs would probably be equally productive, as they are another feature that often carries stance.

A look at students' pronoun usage revealed a tendency to highlight the organizational structure of their texts, explicitly signaling what they are doing or did, for example *I have selected, I will develop, I would like to show*. Hence these are not "author evacuated" (Geertz, 1988), objective, academic texts. The student writers seem to be taking a stance as a "signposter" who helps the reader, rather than as an "expert-analyzer" whose interpretations are accepted as valid. This might be evidence of the developmental stages writers go through, which Pecorari neatly describes: "Learning a skill is rarely a straight line from input to mastery. The novice academic writer must crawl before being able to walk" (2003, p.320). In addition, despite the fact that *we* is frequently used by all the single-author writers, novice and expert, many of those uses are *we* + modal. Many of these occurrences could be replaced by a passive or simply removed without changing the impact of the idea. The *we* + modal combination might be a low-risk strategy for expressing evaluation while at the same time avoiding the more explicitly personal "*I*". The pronoun *we* could allow the individual author to avoid taking responsibility for an idea. Follow-up interview data would provide insight into this strategy.

Text-oriented bundles are another way writers can make the structure of their reasoning explicit without taking an evaluative, authorial stance. Such bundles are frequently used in social sciences texts, where:

> knowledge is typically constructed as plausible reasoning rather than as nature speaking directly through experimental findings . ... text-oriented bundles are heavily used to provide familiar and shorthand ways of engaging with a literature, providing warrants, connecting ideas, directing readers around the text, and specifying limitations (Hyland, 2008, p.16).

It is almost as if the student writers were trying to highlight how logical the sequence of their argument is and how it relates to existing research and theories; the sheer "weight" of several items placed one after the other is considered to be enough to convince the reader of an argument's validity.

Adjectives were explored because it was assumed that they would be used frequently to indicate authorial stance; the data does not confirm this. *New* was often used by NNS students to link works by different authors, without being evaluative. Despite showing awareness of audience, this indicates a reluctance to take risks on the part of these NNS novice writers. Similarly, 12 of the 23 occurrences of *new* in the published texts are embedded in non-evaluative series of words that do not highlight stance. Therefore, neither the native language nor the expertise variable can explain these results. Access to larger corpora might shed light on these findings. Evidence of transfer from the students' native French language was found, especially in the use of the adjectives *present* and *important*. In general, these same students showed more willingness to take risks with adverb and adjective collocates, where they used more types and tokens than the other writers.

The analysis of grading adverbs revealed strategic use of lexical pairing to combine two or more evaluative lexical items (*it may be more useful, the most important from my point of view*). This is another example of students' strategic "stacking" of vocabulary, as if their number would then guarantee the logic of the ideas they express: if varied and lexically-rich conceptual vocabulary is not available, combinations of simpler words might suffice. Only one example of the pairing of such simple words was found in the published texts. The NNS students used proportionately fewer grading adverbs but for both groups of students every occurrence of *very* expressed authorial stance. Lexical pairing may prove to be more prevalent at certain stages in the development of field-specific writing expertise.

4. Conclusion

The corpus-based analysis revealed that the French students' English lexis was quite varied for adverb and adjective collocates, but less accurate and less idiomatic than hoped. The analysis also showed that both NS and NNS students are aware of the need to guide readers through their text, but that they are not yet ready to take on the status of the field expert who evaluates others' work.

The study raises questions about how to analyze learner corpora, as absence of an expression does not mean absence of an idea or function. Automatically generated N-grams and concordances can be useful here, in that they can bring to light expressions in non-standard English such as *an important risk, work in continuity about* that may otherwise go unnoticed. It is almost impossible to anticipate such items when drawing up search lists.

The present research confirms that small corpora of target texts are informative starting points. However, future work will involve increasing the sizes of both the learner corpus and the corpus of published psychology articles, in order to further test some of the questions raised here. In order to better understand the interaction between expertise and formulaic language it would be interesting to further analyze the NNS psychology students' corpus for three types of bundle: research-oriented, text-oriented and participant-oriented. Larger corpora are necessary in order to generate statistically valid results that teachers can apply when designing teaching materials for general EAP or psychology students.

References

Anthony, L. (2007). AntConc (Version 3.2.1.). Retrieved June 19, 2007, from http://www.antlab.sci.waseda.ac.jp/software.html

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Boch, F., Grossmann, F., & Rinck, F. (in press). Le cadrage théorique dans l'article scientifique: un lieu propice à la circulation des discours [Theoretical framing in the research article: a location where discourse circulates]. In S. Marnette (Ed.), *Circulation des discours et liens sociaux: le discours rapporté comme pratique sociale, Actes du colloque international Cit-dit, octobre 5-7 2006* [Proceedings of the Annual Cit-dit Conference, Québec, October 2006]. Québec: NotaBene.

British Academic Written English (BAWE) Corpus. Retrieved February 15, 2009, from http://ota.ahds.ac.uk/headers/2539.xml

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, *23*, 397-423.

Dressen-Hammouda, D. (2008). From novice to disciplinary expert: Disciplinary identity and genre mastery. *English for Specific Purposes*, *27*, 233-252.

Flottum, K., Dahl, T., & Kinn, T. (2006). *Academic voices: Across languages and disciplines*. Amsterdam: John Benjamins Publishing Co.

Geertz, C. (1988). *Words and Lives: The Anthropologist as Author*. Palo Alto, California: Stanford University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. & Sinclair, J. (1999). A local grammar of evaluation. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 74-101). Oxford: Oxford University Press.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, *34*, 1091-1112.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*, 4-21.

Jones, M. & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: an exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 269-291). Amsterdam: John Benjamins Publishing Co.

Meunier, F. & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins Publishing Co.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Co.

Oxford English Reference Dictionary (2nd ed.). (1996). Oxford: Oxford University Press. <oXygen/> XML editor (Version 10). Retrieved July 23, 2007, from http://www.oxygenxml.com/

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, *12*, 317-345.

Schmitt, N., & Carter, R. (2004). *Formulaic sequences: acquisition, processing and use*. Amsterdam: John Benjamins Publishing Co.

Scientext: Un corpus et des outils pour étudier le positionnement et le raisonnement dans les écrits scientifiques. Retrieved April 9, 2010, from http://scientext.msh-alpes.fr/scientext-site/spip.php?article1

Swales, J. M. (1990). *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

In Memoriam

Robert Barr passed away in the early hours of April 9, 2010, following a massive heart attack. Only 35, he was much loved by innumerable friends all over the world and by colleagues at the different French universities where he had worked since his arrival in Savoie in 2002. His kindness, patience, computing genius and amazing culinary skills will be dearly missed. Our thoughts go to his wife, Eva Esposito-Barr, and their soon-to-be-born first child, his family and his in-laws.

Acknowledgments

Thanks to Agnès Tutin, Françoise Boch, Benet Vincent and several colleagues in the UdS Psychology Department for helpful input and discussion.

APPENDIX A

Psychology 3rd Year, Semester 5: Individual Written Project

HANDWRITTEN or LATE work will NOT be accepted.

You will write the introduction to a research article. This section is also called the literature review. In it, you show how your work "fits" into existing work. You will refer to published research.

You will give your teacher one A4 page, printed on both sides, double-spaced, size 12 font; your bibliography should be on another page. Remember to provide an electronic copy (on a disk, by e-mail, etc.) as well.

Step A) Where does the author stand: Continuity or Discontinuity?

In the French text below, the author expresses his/her position in relation to existing work in the field. Circle or underline these expressions.

Continuity: The author shows that their work is an "extension" of existing work.

Discontinuity: The author shows that their work is a "break" from existing work.

> « Toute volonté d'offense mise à part, cette analyze me paraît aussi tortueuse que peu convaincante. C'est donc une solution un peu différente que je voudrais proposer et qui recoupe d'ailleurs certaines analyses que D.L. a menées. J'abandonnerai pour ce faire l'hypothèse trop forte à mon goût d'un N résolument résultatif, pour l'hypothèse d'une même fonction de X dans les deux structures en N et en tout(e) N. Je conserverai l'idée d'une norme sous-jacente à ces constructions, elle sera spécifiée plus avant. »

Find the introduction of a research article published in English in your field. Look for similar expressions. List them and bring them to class.

Step B) Focus on verbs of position

Where does the author **stand** in these places? Read carefully …

I believe = j'ai raison

He who believes = les autres ont tort

Find 4-6 psychology research articles published in English on a subject of your choice (preferably on a subject that you know well or want to know better). Look at the verbs in the introduction sections. Focus on the verbs which express the author's position, for example: *argue, assert, assume, believe*, etc. List these verbs and bring them to class.

Step C) Reconstructing

Take notes from these articles. Remember to record the bibliographic details: author, title, year, journal, page numbers, etc. You will need these notes to write your project. The goal is to integrate these ideas into a coherent text (introduction, body, conclusion). Use linking words, verbs, adverbs and expressions which show **where you stand** on the issue.

Remember to insert the appropriate in-text citations [for example, *(Smith, 1990, 45)*]. You learned how to do this in your methodology classes.

UNDERLINE words and ideas which come from other people, even if you have paraphrased.

Finally, give your work an appropriate title.