



**HAL**  
open science

# Competing influences: the impact of mode and language on verb type and density in French and English scientific discourse

Clive E. Hamilton, Shirley Carter-Thomas

## ► To cite this version:

Clive E. Hamilton, Shirley Carter-Thomas. Competing influences: the impact of mode and language on verb type and density in French and English scientific discourse. *Chimera: Romance Corpora and Linguistic Studies*, 2017, Interdisciplinary and interlinguistic perspectives on Academic Discourse: the mode variable, 4 (1), pp.13 - 34. 10.15366/chimera2017.4.1.001 . hal-01570889

**HAL Id: hal-01570889**

**<https://hal.science/hal-01570889v1>**

Submitted on 1 Aug 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Competing influences: the impact of mode and language on verb type and density in French and English scientific discourse

Clive Hamilton<sup>°</sup>, Shirley Carter-Thomas\*

<sup>°</sup> Université Paris Diderot, CLILLAC-ARP & LATTICE

\*Institut Mines-Télécom (TEM) & LATTICE (UMR 8094), CNRS, ENS PSL Research University, Université Paris 3-Sorbonne Nouvelle

This paper proposes a contrastive analysis of lexical verbs in English and French academic discourse. The EIIDA corpus enables us to adopt a doubly contrastive viewpoint, comparing the frequency and types of verbs used both in research articles and oral presentations in the two languages. The results indicate that mode differences are less distinct in French and that overall lexical and verbal density is higher in English (46% the English oral mode and 61% in the written mode, whereas it is 42% and 47% in the French oral and written mode). In contrast, the French articles and presentations sub-corpora display greater lexical variation, suggesting that information is densely packed in English and the same lexical items are often reused, whereas information is less dense in French but is accompanied by greater lexical diversity and less reuse.

**Keywords:** written and oral academic discourse, lexical density, lexical variation, verbal density

## 1. Introduction

Linguistic and stylistic variation in scientific discourse can be considered from a variety of angles. Previous studies have focused predominantly on the written mode and have examined a wide range of traits believed to be characteristic in a particular genre: for example, rhetorical features (Clyne, 1987), author positioning or stance (Swales & Feak, 2012), lexical bundles and other multiword units (Cortes, 2004; Chen & Baker, 2010; Qin, 2014). Others have chosen lexicogrammatical features such as the use of pronouns (Carter-Thomas & Chambers, 2012; Hartwell & Jacques, 2014), modals (Vázquez, 2010; Carrió-Pastor, 2014)

and verbal processes (Ignatieva & Rodríguez-Vergara, 2015; Nesi & Holmes, 2010) but few have addressed the issue of verb typology and polyfunctionality, except for those investigating the well-documented phenomenon of reporting verbs (e.g., Hyland, 2002a).

While research has also frequently been conducted from a contrastive viewpoint, comparing different disciplines (Hyland, 2000; Fløttum et al., 2006), genres (Biber et al., 2002) or productions by students and experts (Cortes, 2004; Qin, 2014), considerably fewer studies have been devoted to comparisons between spoken and written academic discourse. This is surprising in that research on academic registers has found that mode differences are extremely important in accounting for linguistic variation (e.g. Biber, 2006; Swales, 2004): for instance, spoken academic genres, such as conference presentations, display a very different range of lexico-grammatical features from written academic genres (Carter-Thomas & Rowley-Jolivet, 2001). Our study will therefore further explore these questions of mode differences, comparing conference presentations with research articles in the same disciplines. The second variable examined will be that of language. As several past studies have shown, the process of knowledge transfer in the academic context can be influenced by language and by culture-specific norms conveyed by traditions and the educational system (Bennett, 2010; Fløttum, 2003; Molino, 2011). Once again however these studies have mainly focused on the conventions of academic writing, rather than on spoken academic discourse. The objective of the article is to shed light on the interplay between the two variables: mode (written or oral) and language (French or English)

The EIIDA<sup>1</sup> corpus will enable us to adopt this doubly contrastive approach. Keeping the disciplinary bias stable, we will compare the frequency and types of lexical verb used in research articles and oral presentations in the two languages. Our main hypothesis and research questions (RQ) are set out below.

Hypothesis: As verb choices achieve different communicative effects, we expect the distribution to vary according to genre and mode – the selection of linguistic form being determined by not only what the researcher wishes to say but also how it should be said according to genre conventions.

RQ 1: Do French and English differentiate between the spoken and written modes to the same extent and in the same way?

RQ 2: To what extent are certain verbs exclusive to the written or oral mode?

RQ 3: Which of the four sub-corpora exhibits the greatest variety of verbs?

---

1 See Carter-Thomas & Jacques (this volume) for a description of the rationale and design of the EIIDA corpus.

Identifying and describing these potential differences can prove beneficial to those required to communicate in an intercultural academic setting and to those interested in the intercultural facets or nuances that are part of scientific discourse. Such awareness can be usefully reinvested in second language curriculum design by teachers whose students aim to integrate new academic discourse communities (Hyland, 2008; Woodrow, 2006; Xing et al. 2008) or may allow specialised translators to better apprehend the linguistic reality of different discourse communities, by translating not only the intended meaning of the source text but by appropriately adjusting the discourse to the expectations of the intended audience (Bassnett, 2005; Peterlin, 2008).

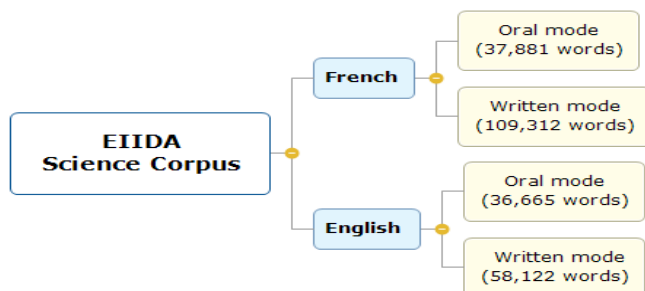
The article is organised as follows. In section 2 we briefly describe the four sub-corpora, and outline some initial hypotheses arising from this description. In section 3 we describe the methodology adopted for analysing the data. Section 4 and 5 offer an account and discussion of the results obtained. Finally, some concluding remarks are made in section 6.

## **2. Corpus characteristics**

The data under investigation are part of the EIIDA corpus, which is divided equally between science (chemistry, geochemistry, marine and water sciences) and the humanities (linguistics) in English, French and Spanish. The current study focuses only on the science corpus in the first two languages. The science subset comprises 30 articles and 30 transcripts of oral conference presentations, which were transcribed using Transcriber (Barras et al. 2001). The 15 English talks and articles correspond very closely. In each case the conference speaker is one of the authors of the corresponding articles. Eleven of the talks and corresponding proceedings articles are taken from the 10<sup>th</sup> Novatech conference, an international conference focusing on water management and related geological and environmental issues. The remaining four articles and corresponding presentations deal with plasma chemistry and oceanography. The French talks were transcribed from recordings made at a conference organised by FROG (French Researchers in Organic Geochemistry). The French science articles were extracted from the journal *Quaternaire*, a peer-reviewed geology journal publishing articles in French and English. Although not an exact match, the French articles were selected so as to be as close as possible thematically to the French conference talks, as well as to the two English sub-corpora.

The talks are relatively short (12–15 minutes), amounting to 36,665 words in the English subset and 37,881 words in the French sub-corpus, while the word count for the articles is considerably higher (58,122 words in the English subset and 109,312 words for the French subset, after anonymization and removal of Acknowledgments and References). Abstracts that were in a different language from the main text were also discarded from the articles sub-corpora. The French articles subset is peer-reviewed and published in a journal, whereas the English articles are conference proceedings – which may partly explain the difference in length.

To allow for the difference in size of the subsets, the results will be expressed either as percentages or normalized to occurrences per 1000 words, to enable comparison between the spoken and written versions. In addition, examples from the written subsets will be indexed by FR-S-E or EN-S-E for French and English respectively, and those from the spoken subset by FR-S-O and EN-S-O, followed by a number referring to the specific presentation or article. Mother tongue status was also verified during the corpus construction to reduce potential second language bias, and to the best of our knowledge all participants are mother tongue speakers in their respective subsets.



**Figure 1.** Overview of the EIIDA science corpus for French and English

### 3. Study design

The first step of the analysis involved the automatic part-of-speech tagging with SketchEnglish and UAM CorpusTool<sup>2</sup> in all four subsets, followed by the ex-

<sup>2</sup> The four sub-corpora were analysed using AntConc (Anthony, 2014), and the integrated version of treetagger in SketchEngine (Kilgariff et al., 2014) and UAM CorpusTool (O'Donnell 2010). Although most of the analysis could have been done with either of the latter, they were both used to keep potential noise to a minimum and ensure reliability. However, only the AntConc figures (obtained using the “latin1” encoding to take into account the French language sub-corpora) were retained for word count.

traction of all the verb-tagged occurrences. The concordances obtained were then examined in context and manually post-edited to discard irrelevant items: for instance, all morphological derivations of verbs were removed (i.e. verb nominalisations, verb-to-adjective derivations and various lexicalised expressions, see examples 1-3 below). This allowed us to focus on those elements that function exclusively as lexical verbs in context. For the same reason, auxiliary and modal verbs were also removed from the initial concordance lines. However, tagged infinitives were included.

- (1) The combination of a directed acyclic (tree-type) modelling environment with a network linear program shows great promise for the realistic *modelling* of cluster size water harvesting, recycling and supply networks. (EN-S-E-09)
- (2) Indeed, most plots of plant height versus hydraulic conductivity at a *given* time showed a negative relationship between the two variables [...] (EN-S-E-04)
- (3) [...] qu'il a corrélié cette occurrence de miliacine dans les sédiments avec la mise en culture du millet sur les versants, autrement *dit* c'est un indicateur de certaines pratiques culturelles. (FR-S-E-14)

The second step of the analysis consisted of ranking and identifying the main verbs according to their frequency. The 20 most frequent lexical verbs from each of the four subsets were then extracted in order to be annotated manually. The verb *show*, for example, was the most frequent with 186 attested occurrences in the English science written mode. The 186 occurrences were sorted into concordance lines and analysed individually. This step was repeated for the 19 remaining most frequent verbs in this English science corpus and again for the three remaining sub-corpora. The ranking obtained is illustrated in Figure 3, Section 4.2.

The third step and most central part of our analysis consisted of examining the selected verb occurrences and analysing the manner and context in which each lexical verb was used by the researcher in each case. Each verbal occurrence was attributed one of four verb labels, which were adapted from Fløttum et al. (2006) and Hyland (2002b). The labels are defined as follows:

- a. Research verbs “refer to the action or the activities directly related to the research process” (Fløttum et al., 2006: 84)

- b. Discourse verbs “denote either processes involving [...] graphical representation [...] or processes directly related to the text structuring and the guiding of the reader” (ibid.)
- c. Argument verbs or position verbs “denote processes related to position and stance, explicit argumentation concerning approval, promotion or rejection” (ibid.)
- d. A fourth category of mental verbs was also adopted. This category denotes the cognitive processes mentioned by the researcher(s).

Although both Hyland (2002b) and Fløttum et al. (2006) suggest that certain verbs that are characteristic of a particular category (for example “*assay, explore, plot, recover...*” are considered research verbs (Hyland 2002a:118), whereas “*illustrate, outline, show ...*” are discourse verbs (Fløttum et al. 2006:84), it is difficult in practice to apply the labels out of context. The semantico-pragmatic roles of verbs can be heavily context-dependent and may consequently vary greatly based on the researcher’s intended meaning. For instance, although the verb *indicate* is often a discourse verb, as in example (4) below, it can also express the author’s position on a given subject (particularly when combined with *would*) in which case we classify<sup>3</sup> it as an argument verb (see example 5).

- (4) The vertical solid line *indicates* the guidelines for human health and the vertical dotted line indicates the [...] (EN-S-E-07)
- (5) This would *indicate* a serious eutrophication potential if these effluents were to be released into [...] (EN-S-E-15)

## 4. Syntactic annotation and verb frequencies

### 4.1 Lexical and verbal density

Previous studies (Biber et al., 1999; Rowley-Jolivet 1998) have shown that lexical density is greater in written than in oral academic discourse: viz. there are more content words than words establishing grammatical relationships (i.e. function or grammatical words). The present study has corroborated those results. Using Ure’s 1971 formula, as reported and adapted by Johansson (2008), content words were restricted to verbs, nouns, adjectives and non-grammatical-

---

<sup>3</sup> A dual-rating system was adopted throughout the classification process until agreement between the two analysts was reached. (See also Jacques & Poibeau, 2010).

ised adverbs. The calculation of lexical density can therefore be schematised as follows:

$$\text{lexicalDensity} = \frac{\text{Lexical words}}{n(\text{words})} \times 100$$

**Equation 1:** Lexical density

The lexical density amounts to 42% in the French oral mode and 47% in the written mode, whereas it is 46% and 61% in the English oral and written mode respectively. The remaining percentages in all four subsets are to be considered as function or grammatical words. The dissimilarity in the range of lexical density can be explained in terms of information packaging, where “a text with a high proportion of content words contains more information than a text with a high proportion of function words” (Johansson, 2008:65). We can therefore assume that the English sub-corpora are more densely packed.

The lack of significant difference in lexical density in the two French sub-corpora allows us to answer one of our initial research questions, i.e. whether the modes can be differentiated to the same extent in both languages. Our preliminary results suggest that there is little distinction between both modes in academic French. In contrast, English seems to display greater variability between written and oral modes. This hypothesis is reinforced further if we consider verb density and verb types in both languages.

Table 1 presents a general overview of verbal occurrences, with the total number of tokens and the initial number of tagged verb occurrences (i.e.  $VB\ n(tags)$ ) and the corresponding percentage ( $VB\ (\%)$ ).

**Table 1.** Overview of verbal occurrences

	$n\ (tokens)$	$VB\ n(tags)$	$VB\ (\%)$
EN-S-O	36665	6514	17.7
EN-S-E	58122	9065	15.5
FR-S-O	37881	4619	12.1
FR-S-E	109312	11110	10.1

Table 2 shows the breakdown into auxiliary and modal verbs (*aux* VB) and full content/lexical verbs (*lx* VB).

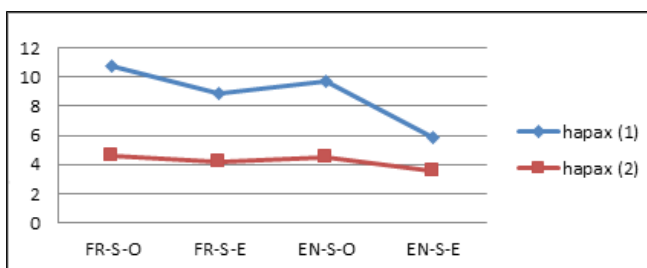


**Table 2.** Breakdown of VB n(tags)

	<b>aux VB (%)</b>	<b>/x VB (%)</b>
EN-S-O	9	8.7
EN-S-E	5.8	9.7
FR-S-O	6.2	5.9
FR-S-E	3.2	6.8

The overall verbal density (i.e. including auxiliaries) for each subset amounts to 12.1% and 10.1% in the French oral and written corpus and 17.7% and 15.5% in the English corpus respectively. In both languages there is a slight difference between the oral and written modes, with a higher density of verbs in the oral presentations than in the articles corpus. In general, speakers use more verbs than writers, segmenting the discourse into smaller chunks which can more easily be processed by listeners. These results corroborate similar patterns identified in oral and written mode comparisons by Biber et al (1999) and Carter-Thomas & Rowley-Jolivet (1998).

Calculating the verbal density has also led however to what might be considered as a counter intuitive observation. The French sub-corpora overall contain a larger number of tokens than the two English sub-corpora, but this is not really significant in terms of either lexical or verbal density. Quite simply, French clauses are longer. Alternatively, one may attempt to explain the numerical difference (cf. number of tokens in Table 1) between the French and English corpus by referring to *hapax legomenon*: i.e. the frequency of words occurring only once in a specific corpus. Figure 2 below provides an illustration of this phenomenon, which was calculated using an example from Linnarud (1976:44). The rationale behind Linnarud's variation count is to supplement lexical density analysis, which cannot alone account for the full range of lexical variation or "lexical variety of a text".

**Figure 2.** Lexical variation in the four subsets

The following formulae were used to calculate lexical variation:

$$VC = \frac{\text{lexicalword hapax}}{n(\text{lexicalwords})} \times 100$$

**Equation 1:** For hapax (1)

$$VC = \frac{\text{lexicalword hapax}}{n(\text{words})} \times 100$$

**Equation 2:** For hapax (2)

The blue line represents the percentage of lexical words that appear only once in each subset in relation to the lexical words that were tagged in that specific sub-corpus: for instance, the noun “sténothermes” was used only once in the entire corpus of written French and is considered a *hapax legomenon* in that subset. The red line indicates the same phenomenon but in relation to the overall number of tokens identified in each sub-corpus. We see therefore that when contrasted with the lexical density percentages mentioned above, Figure 2 shows that the French academic modes tend to have a slightly superior lexical variation count (in both hapaxes), in spite of the fact that lexical and verbal densities are higher in English. In other words, information is densely packed in English and the same lexical items are often reused, whereas information is less dense in French but is accompanied by greater lexical diversity and less reuse.

#### 4.2 Top 20 lexical verbs

Figure 3 below shows the 20 most frequently occurring lexical verbs in the four sub-sets. There are some striking similarities. For example, there are nine verbs in English (*see, look, think, show, find, say, know, put, give*) with their direct equivalents in French (*voir, regarder, penser, montrer, trouver, dire, savoir, mettre, donner*) that appear to be very productive in both oral presentation sub-sets. Similarly, in the written subsets, five verbs appear frequently in both languages (*show, indicate, develop, make, represent* in English, and *montrer, indiquer, développer, faire, représenter* in French). These similarities suggest that certain verbs can be considered characteristic of scientific discourse (within the disciplines examined) and are not language-specific.

ES-S-O			ES-S-E			FR-S-O			FR-S-E		
	<i>lx verb</i>	<i>n</i>		<i>lx verb</i>	<i>n</i>		<i>lx verb</i>	<i>n</i>		<i>lx verb</i>	<i>n</i>
1	see	187	1	show	186	1	voir	136	1	montrer	128
2	look	112	2	model	95	2	dire	59	2	associer	125
3	think	87	3	include	87	3	retrouver	52	3	présenter	124
4	show	79	4	reduce	85	4	montrer	52	4	correspondre	119
5	take	63	5	provide	85	5	utiliser	49	5	faire	92
6	come	57	6	base	78	6	provenir	44	6	observer	91
7	work	55	7	occur	70	7	savoir	39	7	mettre	89
8	find	52	8	increase	63	8	regarder	37	8	marquer	87
9	say	51	9	indicate	53	9	analyser	34	9	dater	86
10	know	43	10	result	52	10	partir	32	10	représenter	83
11	try	40	11	give	51	11	identifier	30	11	lier	82
12	put	39	12	require	51	12	étudier	28	12	caractériser	80
13	start	37	13	develop	50	13	présenter	26	13	indiquer	79
14	want	36	14	consider	50	14	mettre	25	14	témoigner	76
15	talk	34	15	mix	50	15	trouver	25	15	constituer	74
16	give	33	16	make	48	16	observer	23	16	développer	68
17	base	27	17	produce	47	17	obtenir	21	17	réaliser	67
18	make	27	18	find	44	18	penser	21	18	connaître	66
19	mix	24	19	represent	42	19	donner	20	19	situer	66
20	reduce	23	20	apply	37	20	produire	20	20	apparaître	63

**Figure 3.** Most frequent verbs in all four subsets

We also compared the top 20 lexical verbs in the English articles corpus (ES-S-E) with the results of Williams (1996), who examined lexical verbs in two types of medical research reports. Although the studies are not comparable in terms of discipline and text genre, both can be said to belong to a formal written scientific register. In spite of the disciplinary difference, seven of the most frequent verbs (show, find, develop, increase, indicate, include, occur) in our English subset appear to share a similar frequency in Williams' study, suggesting that the frequency in use of some lexical verbs may be similar across different disciplines while others appear to be discipline-specific. Our French verb lists were compared with one established by Yan (2012), who examined the twenty most frequent verbs in a humanities corpus. Although, the corpus used was largely interdisciplinary, 8 of the verbs in FR-S-O (i.e. *voir*, *dire*, *montrer*, *savoir*, *regarder*, *mettre*, *trouver*, *penser*, *donner*) and 7 (i.e. *montrer*, *présenter*, *correspondre*, *représenter*, *lier*, *constituer*, *connaître*) of the verbs in FR-S-E appear in Yan's list, suggesting that these particular verbs may not be discipline specific. Overall there are fifteen verbs that appear in our French list and Yan's study.

Another calculation was made to obtain a normalised figure for the top 20 verbs, i.e. occurrences per 1000 words. This figure was then compared with the previously established lexical variation. The normalised figures (see Table 3) confirm the lexical variation, in that the reuse of frequent verbs is greater in the two English subsets than in the French subsets. For instance, as Table 3 shows,

the most frequent verbs are reused on average 30 times per 1000 words in the oral English mode, while only 20 times in the French oral subset.

**Table 3.** Verbal recurrence

<i>normalised p 1000</i>	
EN-S-O	30.1
EN-S-E	22.7
FR-S-O	20.4
FR-S-E	16.6

These observations on lexical variation can have important implications for second language teaching. For instance, as frequent verbal reuse appears to be a characteristic trait of academic English and less common in academic French, going against these tendencies may prove counterproductive for anyone wishing to engage in knowledge transfer in either language. Not only is verbal frequency important but, given the relative reuse of verbs (22.7% in EN-S-E and 16.6% in FR-S-E), knowledge of specific lexical verbs is a key factor to be taken into consideration. In other words, there may be a real need for a “restricted discipline-based lexical repertoire” for second language learners, as suggested by Hyland & Tse (2007: 235). In this respect, it is also important for French researchers having to communicate their research in English to understand how their own language differs from English. Awareness of these patterns could as a result help both novice first and second language writers to integrate academic discourse communities.

An interesting subsidiary question that arises from this observation is how a discipline specific academic word list should be presented: i.e. in terms of lexical word class or in terms of word families following the example of the Academic Word List (Coxhead, 2000). Although the Academic Word List is undeniably valuable, its usefulness also appears somewhat diluted due to its very broad academic communicative scope. If certain lexical verbs are more important than others, notably in terms of observed frequency in a particular genre or mode, then it would be useful if they were readily identifiable/available. This would increase the readability and applicability of such lists, by instructors and learners alike.

## 5. Semantico-pragmatic annotation and verb types

### 5.1 Polyfunctionality of verbs

The semantico-pragmatic annotation of the 20 most frequently occurring verbs in each subset allowed us to identify another seemingly characteristic feature of academic discourse: the polyfunctional nature of verbs used in oral and written academic French (and to a lesser extent in oral academic English) and the tendency towards verbal monosemy in written academic discourse in English. This observation is based on the individual labelling of our 5018 verbal occurrences (i.e. the sum of each repeated item, as shown in Table 3), where each verb was attributed one of the four verb types mentioned in section 3: *research*, *discourse*, *mental* or *argument*.

Although each individual occurrence was only attributed one label, our results show that some repeated occurrences of the same verb required a different labelling. Three examples of the verb *see* annotated differently are provided below.

- (6) [...] there are unexpected flow patterns in ponds as we've *seen* through several of these presentations [...] (EN-S-O-01)
- (7) [...] in general we get a breakthrough after somewhere around uh depending how you *see* it two and a half thousand and four thousand [...] (EN-S-O-07)
- (8) I'd like to give you some macroscopic effects and that was the whole purpose *to see* how are the evaporation euh rates affected by the various operating conditions of the system [...] (EN-S-O-14)

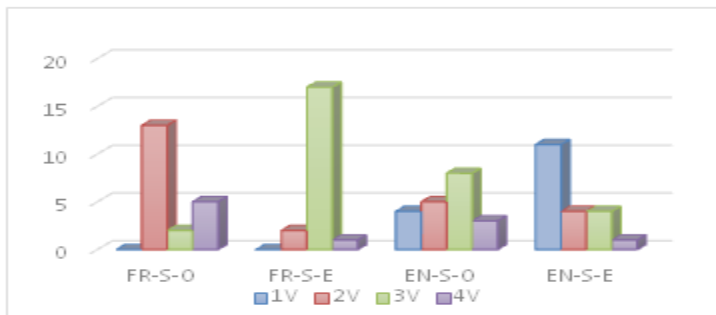
The first occurrence is discursive in nature and can be considered text deixis, in that it refers to earlier segments of discourse or, in this specific case, to earlier presentation segments. The second example was labelled a mental verb, as it refers to the listener's thought processes. The verb *see* can be interchanged here with *consider* or *perceive*. The third example of *see* can be readily replaced by *study* or *examine*, which highlights the author's active role as a researcher. Similar examples from the French subset are presented below.

- (9) [...] donc comme vous pouvez le *voir* sur le gel électrophorèse ici [...] (FR-S-O-15)

- (10) [...] quand on regarde plus précisément le L un on *voit* que à la limite paléocène éocène on a une chute importante de la matière organique gélifiée [...] (FR-S-O-06)
- (11) [...] notre second objectif a été de de *voir* quelle était la diversité moléculaire de ces composés [...] (FR-S-O-15)

Example (9) can be attributed a discourse label, as it suggests a context-dependent spatial reference. Example (10) can be interpreted as a mental verb, as it applies to the cognitive or visual capacity of the listener. The last example can be labelled a research verb, as it shows the speaker’s direct involvement in the research process. The verb *voir*<sup>4</sup> (“see”) can be rephrased as “to determine” or “to get a better understanding”.

This plurality of meaning potential in the annotated verbs suggests that abundant contextual clues are required to facilitate comprehension. Given this factor, such variety of contextual meaning seems better adapted to the oral communication mode where speakers can rely on multiple linguistic and extra-linguistic clues to ensure knowledge transfer: for instance, situational deixis, visual aids and other types of nonverbal communication. If this hypothesis holds true, the polyfunctionality of verbs and the greater lexical variation may be contributing factors to the longer sentences and the overall length of the French subsets, and particularly the written mode.



**Figure 4.** Polyfunctionality of verbs in the four subsets

<sup>4</sup> For more information on the different uses of the verb *voir* in French, see Grossmann (2014).

Figure 4 illustrates the observed reuse of verbs with multiple contrasting semantico-pragmatic and discourse meaning in the four subsets. *1V* represents verbs whose contextual meaning remained unchanged throughout a given subset of texts; *2V* represents verbs with two attested meanings or labels from the four verb types mentioned above, and so forth. What is particularly noteworthy here is that in both French subsets none of the verbs examined were reused in exactly the same manner, or with the same primary meaning, throughout the given subset. In addition, the majority of the verbs (i.e. 15 out of 20 in FR-S-O; and 19 out of 20 in FR-S-E) were attributed two or three labels both in the oral and written modes, suggesting a considerable variety of contextual meaning and indirectly a display of linguistic or stylistic mastery of the language. These semantic variations lend support to Siepmann's claim in which Academic French "is seen as being preoccupied with linguistic artistry" (2006:133). In contrast, however, the English sub-corpora display a more moderate degree of contextual variety. In the English articles corpus (EN-S-E) for example, 11 of the 20 most frequent verbs were attributed only one verb label – implying a certain level of semantic consistency and less semantic polyvalence.

## 5.2 Verb typology and mode differences

Another important aspect that emerges from the semantico-pragmatic annotation concerns the characteristic use and frequency of the four verb types. If we consider one of our initial research questions, that is the possible variation between the oral and written modes and the two languages, we can note that there are substantial differences not only in lexical (and verbal) density, but also in the types of verbs used. Our results indicate slight to moderate variation between the oral and written modes in French, whereas there are significant differences in English oral and written scientific discourse

This observation was made after calculating the percentage of each verb type, within a specific subset. For instance, in the French oral corpus there are 97 annotated examples of verbs denoting cognitive processes that were tagged as mental verbs. This figure represents 13.4% of all the tagged items, when compared to the remaining three verb types that were used in that subset (see Table 5). Once this calculation was completed for all four subsets, the data was compared between modes (i.e. between FR-S-O and FR-S-E; and between EN-S-O and EN-S-E) and then between languages.

A rating scale is provided to better illustrate the comparison that follows. For instance, when the difference (D) in percentage between any two groups un-

der comparison is between 0% and 5%, it is considered unimportant; when (D) varies from 6% to 10%, it is considered somewhat important, and so forth.

**Table 4.** Rating scale

Rate (%)	Difference
0 → 5	unimportant
6 → 10	somewhat important
11 → 15	moderately important
16 → 20	very important
20 → 25	extremely important

When we consider the variations between the two languages in oral scientific discourse, some small differences emerge (see Table 5). In the English conference presentations, discourse verbs (often used for signalling the structure of the presentation, as in (12) below) are more frequent than in the French talks, whereas argument verbs are more prevalent in the French subset. However, research and mental verbs are used in a strikingly similar manner in both languages. In general, there appears to be only slight to moderate variation in verb use, suggesting perhaps that in the scientific disciplines examined French and English oral academic discourse share more commonalities than dissimilarities.

- (12) [...] but of course they also provide quite significant water quality benefits and *I'll talk* a little bit about that in a minute [...] (EN-S-O-07)

**Table 5.** Variation in French and English oral discourse

	EN-S-O	FR-S-O	(D)	
Mental	12.5	13.4	0.9	unimportant
Discourse	28.3	12.6	15.7	moderately important
Argument	6.9	21.2	14.3	moderately important
research	52.1	52	0.1	unimportant

Written scientific discourse (see Table 6), on the contrary, does not display the same level of convergence. There are notable differences in the use of argument and research verbs. For example, there are considerably less research verbs in the French subset than in the English subset, and the French writers seem to compensate by having recourse to more argument verbs. It is noteworthy here that argument verbs are practically absent in the written mode in English, where



there seems to be a marked preference for research and discourse verbs – which account for 92% of the semantico-pragmatic annotations in this subset. These patterns of verb use in the written mode raise the question as to whether recourse to more argument verbs in the FR-S-E corpus (as in (13) below) indicates heightened author positioning vis-à-vis the discourse content, which inevitably results in more argumentation in support of claims, or whether the research and discourse verbs combination (EN-S-E corpus) suggests a focus on content and signposting rather than author positioning, as in (14).

- (13) *Pourtant, l'anthracologie montre que Quercus ilex, Arbutus unedo, Phillyrea/Rhamnus et Pistacia, sont les essences dominantes dans le bois de feu des populations du Bronze final à La Fangade.* (FR-S-E-07)
- (14) The colour shading shown in Figure 1.1 *represents* depth (in metres) beneath the permanent water surface elevation and the secondary grey contour lines *represent* 0.5 m elevation contours. (EN-S-E-01)

**Table 6.** Variation in French and English written discourse

	EN-S-E	FR-S-E	(D)	
Mental	5.7	1.4	4.3	Unimportant
Discourse	13.8	7.7	6.1	somewhat important
Argument	2.3	24.8	22.5	extremely important
Research	78	65.9	12.1	moderately important

The comparison between modes within the same language also suggests some significant differences: on the one hand, the absence of any strong variation between the oral and written modes in French (see Table 7) and on the other hand, the extreme volatility in the way the four verb labels are used in the two modes in English (see Table 8).

**Table 7.** Variations between modes in English

	EN-S-O	EN-S-E	(D)	
Mental	12.5	5.7	6.8	somewhat important
Discourse	28.3	13.8	14.5	moderately important
Argument	6.9	2.3	4.6	unimportant
Research	52.1	78	25.9	extremely important

In English mental, discourse and argument verbs are all used more in the oral mode than in writing. Accounting for 48% of overall annotations in the oral subset, the three verb categories together only represent 22% in the written mode. Research verbs on the contrary show a correspondingly greater use in writing.

**Table 8.** Variations between modes in French

	FR-S-O	FR-S-E	(D)	
Mental	13.4	1.4	12	moderately important
Discourse	12.6	7.7	4.9	unimportant
Argument	21.2	24.8	3.6	unimportant
Research	52	65.9	13.9	moderately important

However, while all four verb categories used appear sensitive to the question of mode in English, only mental (or cognition) verbs also appear sensitive to mode in French. Our results show that mental verbs are infrequent in both French and English written academic discourse (1.4% and 5.7% respectively), whereas they are equally frequent in the English and French oral presentations (12% and 13%). This is undoubtedly because, as opposed to scientific articles in which the focus is on completed research, the emphasis in conference presentations is often on work in progress and new development and mental verbs serve to express the conceptualisation of this ongoing research.

- (15) [...] nous ne *savons* pas exactement par quel moyen cet échantillon  
[...] (FR-S-O-01)
- (16) [...] donc on *pense* que ça pourrait venir de euh des taux de [...] (FR-S-O-08)

This genre preference can be considered to override the language influence here. In French, however, there appears to be little variation in the use of the other verb categories, with argument verbs in particular showing strikingly similar proportions of use in the two modes.

## 6. Concluding remarks

In this article, we have examined the competing influences of mode and language on verb use and typology in academic discourse. The different comparis-

ons have supported the hypothesis that written and oral science make use of different verbal strategies but have also shown that the interplay between mode and language is fairly complex.

Regarding research question n° 1, mode appears to exert a strong influence in the English subsets. Not only are differences in lexical and verbal density more pronounced in the two English subsets, but the differences in the types of verbs used in the two modes are also very marked. French, however, generally appears less sensitive to the question of mode.

Another striking difference between the two languages concerns verbal polyvalence. With regard to research question n° 2, the French verbs analysed in both oral and written subsets displayed a marked degree of contextual variety. They were consistently categorised with at least two or three different semantico-pragmatic labels - out of the four types examined in this study - and used consequently for different purposes. English verbs, in contrast, demonstrated a tendency towards more semantic consistency, often being associated only with one label. This difference is particularly marked in the case of English and French written scientific discourse.

Concerning research question n° 3, French appears very sensitive to the question of language reuse. Lexical variation is more pronounced and reuse is lower. Whereas English relies more on a smaller set of specific verbs where the same items are often reused, French academic discourse appears to use a wider set of verbs with less repetition.

Our results also indicate that knowledge of individual lexical items is an insufficient criterion for studying or evaluating academic discourse. Lexical and verbal density, the degree of lexical and verbal variation and verb typology are all important features to be taken into consideration, especially for novice or second-language researchers wishing to join these academic discourse communities.

The design of the EIIDA corpus has given us the opportunity to focus on the interplay of language and mode variables and suggests several profitable avenues for further research. As however the corpus is relatively small, any overall conclusions should be drawn with caution. In the future, it would be interesting to also include the discipline variable and examine mode and language differences in a larger corpus across multiple disciplines, so as to ensure greater generalisability.

## Acknowledgments

EIIDA (Etude Interlinguistique et Interdisciplinaire des Discours Académiques, *Interlinguistic and Interdisciplinary study of Academic Discourse*) was a 3-year project (2012–2015) funded in the framework of the Labex TransferS programme, Ecole Normale Supérieure, Paris.

## References

- Anthony, L. 2014. AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University.
- Banks, D. 2008. *The Development of Scientific Writing. Linguistic features and historical context*. London: Equinox.
- Barras, C., Geoffrois, E., Wu, Z & Liberman, M. 2001. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication* 33: 5-22.
- Bassnett, S. 2005. Bringing the News Back Home: Strategies of Acculturation and Foreignisation. *Language and Intercultural Communication* 5: 120-130.
- Bennett, K. 2010. Academic discourse in Portugal: a whole different ballgame? *English for Academic Purposes* 9(1): 21-32.
- Biber, D. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited
- Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. 2002. Speaking and writing in the university: a multidimensional comparison. *TESOL Quarterly* 36: 9-48.
- Carrió-Pastor, M. L. 2014. Cross-cultural Variation in the Use of Modal Verbs in Academic English. *SKY Journal of Linguistics* 27: 153-166.
- Carter-Thomas, S. & Rowley-Jolivet, S. 2001. Syntactic differences in oral and written scientific discourse: the role of information structure, *ASP*, 31-33, 19-37.
- Carter-Thomas, S & Chambers, A. 2012. From text to corpus: A contrastive analysis of first person pronouns in economics article introductions in English and French. In A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (eds) *Corpus-Informed Research and Learning in ESP: Issues and applications*. Amsterdam: Benjamin, 15-44.

- Chen, Y. & Baker, P. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14 (2): 30–49.
- Clyne, M. 1987. Cultural differences in the organisation of academic texts: English and German. *Journal of Pragmatics* 11: 211-247
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes* 23: 397– 423.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2): 213-238
- Fløttum, K. 2003. « JE » et le verbe. *Tribune* 14: 7-14.
- Fløttum, K., Dahl, T. & Kinn, T. 2006. *Academic Voices – across languages and disciplines*. Amsterdam: John Benjamins.
- Grossmann, F. 2014. Les verbes de constat dans l'écrit scientifique. In A. Tutin et F. Grossmann. *L'écrit scientifique : du lexique au discours*. Rennes : Presses Universitaires de Rennes, 85-100.
- Hartwell, L. M & Jacques, M-P. 2014. Authorial Presence in French and English: "Pronoun + Verb" Patterns in Biology and Medicine Research Articles. *Discours*, 15. TEI. <http://discours.revues.org/8941> (accessed October 1, 2016).
- Hiltunen, T. 2006. Coming-to-know verbs in research articles in three academic disciplines. *Proceedings of the fifth International AELFE (Asociación Europea de Lenguas para Fines Específicos) Conference*, 246-251.
- Hyland, K. 2000. *Disciplinary discourses: Social interactions in academic writing*. Harlow: Longman.
- Hyland, K. 2002a. Activity and evaluation: reporting practices in academic writing. In J. Flowerdew (ed) *Academic discourse*. London: Longman, 115-30.
- Hyland, K. 2002b. Directives: Argument and Engagement in Academic Writing. *Applied Linguistics* 23(2): 215-239.
- Hyland, K. 2008. Plenary Speeches: Genre and academic writing in the disciplines. *Language Teaching* 41(4): 543–562.
- Hyland, K., & Tse, P 2007. Is there an "Academic Vocabulary"? *TESOL Quarterly* 41(2): 235-253.
- Ignatieva, N & Rodríguez-Vergara, D. 2015. Verbal processes in academic language in Spanish: exploring discourse genres within the systemic functional framework. *Functional Linguistics* 2(2). TEI. <https://functionallinguistics.springeropen.com/articles> (accessed June 1, 2016)
- Jacques, M-P & Thierry, P. 2010. Étudier des structures de discours : préoccupations pratiques et méthodologiques. *CORELA* 2 (8). TEI.

- <http://09.edel.univ-poitiers.fr/corela/index.php?id=1855> (accessed October 1, 2016).
- Johansson, V 2008. Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund: Working Papers* 53: 61-79.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovár, V., Michelfeit, J., Rychlý, P., Suchomel, V. 2014. The Sketch Engine: ten years on. *Lexicography*, 1-30
- Linnarud, M. 1976. *Lexical density and lexical variation - An analysis of the lexical texture of swedish students' written work*. Poznań: Adam Mickiewicz University.
- Molino, A. 2011. A contrastive study of knowledge claims in linguistics research articles in English and Italian. *ESP Across Cultures* 8: 89-101.
- Nesi, H. & Holmes, J. 2010. Verbal and mental processes in academic disciplines. In M. Charles, D. Pecorari & S. Hunston (eds). *Academic Writing: At the Interface of Corpus and Discourse*. London: Continuum, 58-72.
- O'Donnel, M. 2010. UAM CorpusTool (version 3.2.). [Computer Software]. Madrid, Spain: Universidad Autónoma de Madrid.
- Peterlin, A.P. 2008. The thesis statement in translations of academic discourse: an exploratory study. *The Journal of Specialised Translation* 10: 10-22.
- Qin, J. 2014. Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System* 42: 220-231.
- Rowley-Jolivet, E. 1998. *La Communication scientifique orale : Étude des caractéristiques linguistiques et discursives d'un genre*. PhD diss., Université Bordeaux 2.
- Siepmann, D. 2006. Academic writing and culture: An overview of differences between English, French and German. *Meta: Translators' Journal* 51 (1): 131-150.
- Swales, J. 2004. *Research Genres*. Cambridge: Cambridge University Press.
- Swales, J. & Feak, C. 2012. *Academic Writing for Graduate Students: Essential Tasks and Skills, 3rd Edition*. Michigan: University of Michigan Press/ELT.
- Vázquez, I. 2010. A contrastive analysis of the use of modal verbs in the expression of epistemic stance in business management research articles in English and Spanish. *Ibérica* 19: 77-96.
- Williams, I.A. 1996. A contextual study of lexical verbs in two types of medical research reports: clinical and experimental. *English for Specific Purposes* 15(3): 175-197.

- Xing, M., Wang, J. & Spencer, K (2008) Raising students' awareness of cross-cultural contrastive rhetoric in English writing via an e-learning course. *Language Learning & Technology* 12 (2): 71-93.
- Yan, R. 2012. Observation du lexique verbal transdisciplinaire et modélisation des patrons dans l'écrit scientifique: construction d'un dictionnaire électronique d'apprentissage d'encodage. Master diss., Université Stendhal (Grenoble-III).