Lexical bundles and L1 transfer effects*

Magali Paquot
Université catholique de Louvain

This exploratory study makes use of Jarvis's (2000) methodological framework to investigate transfer effects on French EFL learners' use of lexical bundles. The study focuses on 3-word recurrent sequences that include a lexical verb in the French component of the International Corpus of Learner English (ICLE) as compared to nine other ICLE learner sub-corpora. Results are in line with a usage-based view of language that recognizes the active role that the first language (L1) may play in the acquisition of a foreign language. The different manifestations of L1 influence displayed in the learners' idiosyncratic use of lexical bundles are traced back to various properties of French words, including their collocational use, lexico-grammatical patterns, function, discourse conventions, and frequency of use. Following Hoey (2005), these transfer effects are subsumed under the general term of 'transfer of primings'.

Keywords: phraseology, lexical bundles, lexical verbs, transfer, lexical primings

1. Introduction

The last ten years have witnessed a remarkable boom in the number of studies that examine learners' use of lexical bundles, i.e. "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (Biber et al. 1999: 990). These repeated sequences of words may be grammatically complete (*by contrast, on the other hand*) or incomplete (*the nature of the, is based on the*). They may be composed of clause segments (e.g. *I don't know what*) or be parts of phrases (e.g. *the use of*). They are "conventionalized building blocks that are used as convenient routines in language production" (Altenberg 1998: 122) and typically function as referential markers (e.g. *the end of the*), text organizers (e.g. *on the basis of, for example*), stance markers (e.g. *it is possible to*) or interactional discourse markers (e.g. *or something like that, thank you so*

much) (Biber et al. 2003). Corpus-query tools often provide an option that retrieves repeated sequences of words of a given length (e.g. two-word sequences, three-word sequences) fully automatically. This method has been used extensively to compare the number of lexical bundles, their structural characteristics and discourse functions in learner and native corpora.

As pointed out in Paquot & Granger (2012: 138), the results of such studies are particularly difficult to compare. The lexical bundles investigated are of different sizes (from two- to six-word bundles), and the settings used to extract them may vary considerably. However, a number of general trends can be identified. Learners tend to use more lexical bundles in writing when compared to native speakers, but the overall number of recurrent word combinations tends to decrease as proficiency in the language (Reppen 2009) or the time spent in the target language environment (Groom 2009) increases. Most studies report a mixed pattern of under- and overuse. For example, learner writing is often characterized by an underuse of the most academic-like bundles, such as noun phrases with postmodifier fragments (e.g. *the idea that, the issue of*), coupled with an overuse of speech-like word sequences, such as *and so, sort of* and *a lot of* (De Cock 2003, Juknevičienė 2009).

Some of the studies have put specific patterns of misuse, overuse and underuse of lexical bundles down to the learners' mother tongue. Allen (2011: 111), for example, attributes Japanese learners' overuse of *it can be said that* to the L1, as its translational equivalent is repeatedly used in Japanese academic writing. Rica (2010) notes that a large proportion of the multi-word connectors that Spanish EFL writers overuse are very similar to the word sequences used in Spanish to express similar meanings (e.g. En. "I think" and Sp. "Creo que", En. "for example" and Sp. "por ejemplo"). However, no study has targeted transfer effects on EFL learners' production of recurrent word sequences as their primary object of investigation.

The main objective of the present work is to fill in this existing gap by conducting a careful transfer study of lexical bundles in learner writing. The study focuses on 3-word recurrent sequences that include a lexical verb in French EFL learner writing and addresses the following research questions:

(i) RQ1: How much of French learners' idiosyncratic use of lexical bundles with verbs can be attributed to L1 influence?

(ii) RQ2: What type of transfer effect (e.g. transfer of form, transfer of function) is most discernible?

It is hypothesized that lexical bundles are potentially transferable because they are essentially semantically and syntactically compositional, thus typically unmarked word combinations (cf. Kellerman 1978). It is also anticipated that transfer effects will be particularly noticeable in the overuse of lexical bundles whose equivalent forms fulfil specific discourse functions in French.

To answer the research questions, the study is grounded in Jarvis's (2000) unified framework for the study of L1 influence (Section 2). Section 3 describes the learner and native corpus data used. In Section 4, the different methodological steps required are summarised. Section 5 offers the results of the analysis of transfer effects on French EFL learners' use of lexical bundles, and Section 6 provides answers to the research questions in the light of the preceding sections. Section 7 contains concluding remarks.

2. Jarvis's (2000) unified framework for the study of L1 influence

Transfer studies have too often fallen into the trap of making a case for L1 influence on the sole argument that the structure exists in the L1, thus relying exclusively on "'shot-in-the-dark' post hoc interpretive guesses which pass for explanations" (Lightbown 1984: 245). To remedy this situation, Jarvis (2000) put forward a unified framework for the study of L1 influence which is premised on the following operationalizable definition of the construct of 'transfer':

> L1 influence refers to any instance of learner data where a statistically significant correlation (or probability-based relation) is shown to exist between some features of learners' IL performance and their L1 background.
>
> (Jarvis 2000: 252)

This definition of L1 influence translates into a list of at least three potential sources of evidence that transfer studies should consider altogether when presenting a case for or against L1 influence:

(i) **Effect 1: Intra-L1-group homogeneity in learners' IL performance** is found when learners who speak the same first language behave as a group with respect to a specific second language (L2) feature. To illustrate this first L1 effect, Jarvis uses Selinker's (1992) finding according to which Hebrew-speaking learners of English as a group tend to produce sentences in which adverbs are placed before the object (e.g. *I like very much movies*).

(ii) **Effect 2: Inter-L1-group heterogeneity in learners' IL performance** is found when "comparable learners of a common L2 who speak different L1s diverge in their IL performance" (Jarvis 2000: 254). To illustrate, Jarvis (2000) refers to a number of studies reported by Ringbom (1987) that have shown that Finnish-speaking learners are more likely than their Swedish-speaking counterparts to omit English articles and prepositions. Jarvis (2000) argues that "this type of evidence strengthens the argument for L1 influence because it essentially rules out developmental and universal factors as the cause of the observed IL behaviour. In other words, it shows that the IL behaviour in question (omission of function words) is not something that every learner does (to the same degree or in the same way) regardless of L1 background" (Jarvis 2000: 254–255).

(iii) **Effect 3: Intra-L1-group congruity between learners' L1 and IL performance** is found where "learners' use of some L2 feature can be shown to parallel their use of a corresponding L1 feature" (Jarvis 2000: 255). This is the type of evidence that Selinker (1992) produced when showing that Hebrew-speaking learners' positioning of English adverbs parallels their use of adverbs in the L1. The added value of this third

effect is that it also has explanatory power by showing what in the first language motivates the IL behaviour.

In a follow-up article, Jarvis (2010) acknowledges the existence of a fourth type of evidence that was not accounted for in his original framework, viz. 'intralingual contrasts', which he defines as "differences in learners' performance on features of the target language that vary with respect to how they correspond to features of the source language" (Jarvis 2010: 175).

3. Data

The learner corpus data used for the present study come from the first version of the International Corpus of Learner English (ICLE) (Granger et al. 2002). ICLE texts share a number of learner and task variables, which were used as corpus-design criteria. All the learners are young adults who study English as a Foreign Language (EFL) at university. They are all in their second, third, or fourth year and their proficiency level has commonly been described as advanced although learner groups differ in proficiency (Granger et al. 2009: 12). Learner productions share many task variables, notably for medium (writing), genre (academic essay), field (general English rather than English for Specific Purposes) and length (between 500 and 1,000 words). Other variables differ. A majority of the learner texts are argumentative, but the essays cover a wide range of topics (e.g. the death penalty, euthanasia,). Learner texts also differ in task conditions.

The focus of this study is on French learner writing but other ICLE sub-corpora were also used as comparable corpora to test for inter-L1-group heterogeneity in learners' interlanguage (IL) performance (cf. Section 2). Table 1 provides a breakdown of the ten ICLE sub-corpora used. Learner essays in each sub-corpus were carefully selected in an attempt to control for a number of task variables which may affect learner productions (cf. Kroll 1990, Ädel 2008): all the texts are untimed argumentative essays, potentially written with the help of reference tools. Although essays written without the help of reference tools would arguably have been more representative of what advanced

EFL learners can produce, untimed essays with reference tools are used as they represent the majority of learner texts in ICLE.

Table 1. Breakdown of ICLE essays

	No. of essays	**No. of words**	**Average no. of words per essay**
Corpus under analysis			
French (ICLE-FR)	228	136,343	598
Comparable corpora			
Czech (ICLE-CZ)	147	130,768	890
Dutch (ICLE-DU)	196	162,243	828
Finnish (ICLE-FI)	167	125,292	750
German (ICLE-GE)	179	109,556	612
Italian (ICLE-IT)	79	47,739	604
Polish (ICLE-PO)	221	140,521	636
Russian (ICLE-RU)	194	165,937	855
Spanish (ICLE-SP)	149	99,119	665
Swedish (ICLE-SW)	81	48,060	593
TOTAL	**1,641**	**1,165,524**	**697**

To evaluate Effect 3, several corpora of French writing were used. The 1.6 billion word frWaC was consulted via the Sketch Engine (Kilgarriff & Kosem 2012) to assess the frequency and prototypicality of specific word combinations in French for general purposes. It was also deemed necessary to query smaller but more comparable corpora of expert and student writing to control for text type and levels of writing expertise:

(i) The humanities component of the online Scientext corpus, i.e. a 3,431,531 word corpus of French published articles, theses and proceedings in linguistics, psychology, education and natural language processing.

(ii) The Corpus de Dissertations Françaises (CODIF), i.e. a 92,832 word corpus of argumentative essays written by French-speaking students on similar topics to ICLE-FR.

Spot-checks were also sometimes made in the 100 million word British National Corpus and the 2 billion word English corpus ukWaC to check the lexicogrammatical and

distributional properties of English word combinations and hence identify possible intralingual contrasts (Jarvis 2010). The two corpora were queried via the Sketch Engine (Kilgarriff & Kosem 2012).

4. Methodology

The methodology used involves several steps which are described here. Section 4.1 covers the extraction of lexical bundles from ICLE texts. Section 4.2 provides the procedures and statistical tests used to operationalize Jarvis's (2000) unified framework on learner corpus data. Section 4.3 describes the method used to rule out topic influence and the rationale behind this extra step.

4.1 Extraction of lexical bundles

The focus of the study is on potential transfer effects on French EFL learners' use of bundles with lexical verbs. Lexical bundles of 3 words were first extracted from the ICLE French sub-corpus with the help of the computer software WordSmith Tools 5 (Scott 2008). A minimum frequency threshold of 5 occurrences was adopted. The resulting list was filtered manually and the bundles that included a lexical verb were selected for further analysis. A Perl program was then used to retrieve relative frequencies per 100 words for each of the selected bundles in the 1,641 learner texts that make up the ten learner corpora.

4.2 Applying Jarvis's (2000) unified framework to learner corpus data

Intra-L1-group homogeneity is most evident when directly compared with inter-L1-group heterogeneity (Jarvis 2000), and I therefore make use of comparison of means tests to operationalize Jarvis's unified framework for the study of L1 influence on learner corpus data. As more than two learner populations are being compared, one-way

between-groups analysis of variance (ANOVA) tests are used to measure the first two potential L1 effects, i.e. intra-L1-group homogeneity and inter-L1-group heterogeneity. An ANOVA examines two sources of variance: the variance between the groups (i.e. inter-L1-group heterogeneity between the different ICLE sub-corpora) and the variance between individuals or texts within each group (i.e. intra-L1-group homogeneity as displayed in each ICLE sub-corpus). The two types of variance are then compared with one another. If the variance between the learner corpora is significantly higher than the variance within each learner corpus, the interpretation is that the corpora are not taken from the same population. The result of an ANOVA is an F ratio which tells us whether at least one group in the set is different from the other groups. The level of risk or level of significance used in this study is $p < 0.01$.

Importantly, while an F ratio indicates whether a significant difference exists somewhere between the learner populations, it does not identify precisely where the difference is. A post-hoc test must then be conducted to pinpoint the learner population(s) responsible for the significant difference. As the objective here is to evaluate Effects 1 and 2, the comparisons of interest are those between the French learner corpus and the other ICLE sub-corpora. The Dunnett's test is considered the most powerful post-hoc test whenever one group is compared with each of the other groups (Howell 1997: 380-381) and is therefore used in this study.[1] When lexical bundles display significant differences in use between the French learner group and at least half of the other learner populations as revealed by Dunnett's tests, there is a strong case for intra-L1-group homogeneity and inter-L1-group heterogeneity. The criterion used according to which over half of the comparisons need to be significant is arbitrary and probably a relatively conservative estimate. It is, however, used in this exploratory study to validate the methodology. All statistical tests were performed with R (R Core Team 2012).

While the first two effects readily lend themselves to automatic and quantitative evaluation, intra-L1-group congruity between French learners' L1 and IL performance does not. Assessing this third effect requires a more qualitative approach. First, the use of each lexical bundle was carefully analysed in ICLE-FR. The next steps consisted in identifying the French potential "equivalent" of each lexical bundle in context, describing its use in French L1 and comparing learners' L1 and IL patterns of use.

4.3 Addressing the issue of topic variability in ICLE

Learner texts in ICLE are varied in topic, and there is no single topic that is evenly distributed across the 10 sub-corpora used in this study. Topic variability must however be addressed as lexical bundles are particularly prone to this factor (Cortes 2004) and the ICLE French sub-corpus is characterised by a strong bias towards just one topic ("Europe 92: loss of sovereignty or birth of a nation?"). This topic was selected by c. 40% of all the French learners, and more than 70% of all the texts about Europe 92 in ICLE are to be found in the French component. As the issue of topic variability could not be addressed a priori, it is dealt with just before intra-L1-group congruity between French learners' L1 and IL performance (Effect 3) is tested. To rule out topic influence, the ICLE in-built corpus query tool is used to analyse the distribution by essay prompt of all the bundles that display intra-L1-group homogeneity and inter-L1-group heterogeneity (Effects 1 and 2). If a lexical bundle only appears in French learners' essays discussing the creation and future of Europe and in no other ICLE text, this provides a strong indication that topic is a much more likely explanation than L1 influence.

5. Results

This section presents the results obtained from the transfer study. The extraction procedure outlined in Section 4.1 made it possible to identify 273 bundles with a lexical verb in the French learner corpus, which were submitted to further analysis.

5.1 Testing Effects 1 and 2

An R script was written to assess Effects 1 and 2 for the 273 lexical bundles under study. The ANOVA test identified 87 lexical bundles that present significant differences

in use among the ten learner corpora. Among these, 34 bundles (12.45%) display significant differences in use between the French learner group and at least half of the other learner populations as revealed by Dunnett's tests, thus showing both intra-L1-group homogeneity and inter-L1-group heterogeneity. Table 2 lists the 34 bundles, their F ratio and *p* value, as well as the number of learner populations from which the French learner group differs significantly in its use of each lexical bundle.

Table 2. The 34 bundles that show Effects 1 and 2

Bundle	**F**	***p***	**Number of significant learner corpus comparisons**
be considered as	3.075	0.00116	6
be tempted to	4.534	6.45e-06	9
considered as a	4.947	1.4e-06	9
considered as the	2.876	0.00226	7
deeply rooted in	3.101	0.00106	8
does it mean	2.99	0.00154	8
going to become	2.813	0.00278	8
I would say	3.142	0.000919	6
is to know	3.195	0.000767	9
keep its own	3.839	8.03e-05	9
keep their own	3.822	8.54e-05	9
**loose their identity*	2.463	0.00867	7
not forget that	6.457	4.59e-09	9
role to play	2.947	0.00178	9
say that Europe	4.723	3.21e-06	9
speak of a	2.737	0.00357	8
take the example	5.121	7.3e-07	9
to be found	5.206	5.32e-07	7
to build a	4.274	1.67e-05	9
to create a	2.788	0.00302	6
to go further	2.485	0.00809	6
to know whether	2.85	0.00246	8
wait and see	4.699	3.52e-06	9
want to create	3.011	0.00143	8
was considered as	2.421	0.00991	6
we can say	3.192	0.000774	6
we can wonder	2.669	0.00446	6
we may wonder	3.338	0.000469	9
we must not	2.606	0.00549	8
will be allowed	3.261	0.000612	8
will be needed	3.299	0.000536	9
will be united	3.328	0.000484	9
will keep its	3.309	0.000518	9
would say that	3.696	0.000134	8

5.2 The influence of the topic

An analysis of the 34 significant bundles in the 1,641 learner texts and their distribution by essay prompt reveals that 14 lexical bundles only appear in ICLE-FR essays that discuss the creation and future of Europe. These bundles are *keep its own, keep their own, say that Europe, to build a, wait and see, will be needed, will be united, will keep its, will be allowed, does it mean, going to become, want to create, *loose their identity*, and *to create a*. The influence of topic is visible in the selection of content words (e.g. *say that Europe, want to create*) as well as in tense preferences (e.g. *will be allowed, will be united, will keep its*) (Examples (1) and (2)).

(1) Europe **will be united** against USA and Japan. (ICLE-FR)

(2) Each country **will keep its own** identity, currency, institutions and constitution. (ICLE-FR)

The influence of topic was ruled out for the remaining 20 lexical bundles as they were found in essays covering a range of prompts (cf. Table 3).

Table 3. 20 lexical bundles for which topic influence is ruled out (ordered by decreasing frequency in ICLE-FR)

Lexical bundle	Freq.	Rel. freq. (100,000 words)	Texts
we can say	22	16.1	16
I would say	20	14.7	16
would say that	19	13.9	15
not forget that	19	13.9	18
considered as a	18	13.2	17
be considered as	18	13.2	17
to be found	17	12.5	17
we must not	12	8.8	11
take the example	10	7.3	9
considered as the	8	5.9	8
was considered as	7	5.1	6
deeply rooted in	7	5.1	6
be tempted to	7	5.1	7

is to know	7	5.1	6
speak of a	6	4.4	5
to know whether	6	4.4	6
we may wonder	6	4.4	5
we can wonder	5	3.7	5
to go further	5	3.7	5
role to play	5	3.7	5

5.3 Testing Effect 3

The simplest way to test Effect 3 is to check whether there are equivalent lexical bundles in French. Before doing so, however, a quick scan of concordance lines for the 20 remaining lexical bundles (Table 3) showed that some regrouping of embedded word sequences was possible (sometimes making up longer and more syntactically complete bundles such as *I would say that* or pinpointing shorter but more salient word combinations, e.g. *considered as*). Intra-L1-group congruity between learners' L1 and IL performance was consequently evaluated for fifteen lexical bundles (see Table 4). L1/IL equivalence in form was found for a majority of the English lexical bundles; equivalence in meaning or function was established for the four lexical bundles involving the first person plural pronoun *we*. Table 4 also provides the most frequent corresponding bundles in French as identified in frWaC for each of the fifteen longer, syntactically complete or more salient lexical bundles. Small capitals are used to represent lemmas rather than word forms. The extent of the correspondence between the English and French lexical bundles is discussed in Section 6.

Table 4. Lexical bundles and their most frequent equivalent forms in French

English lexical bundles	**Most frequent equivalent bundles in French**
be tempted to	*être tenté/es de*
considered as	*considéré/es comme*
deeply rooted in	*profondément enraciné/es dans*
I would say that	*je dirais que*
is to know whether	*est de savoir si*
not forget that	*pas oublier que*
role to play	*rôle à jouer*
speak of	*PARLER de*
take the example	*PRENDRE l'exemple*
to be found	*être trouvé/es*
to go further	*aller plus loin*

we can say	*on peut dire*
we can wonder	*on peut se demander*
we must not	*il ne faut pas*
we may wonder	*on peut se demander*

6. Discussion

This section addresses the research questions guiding the study by discussing the results provided in Section 5. The combination of the three effects investigated in Section 5 points to a firm conclusion of L1 transfer for the twenty lexical bundles for which topic influence was ruled out (Table 3). This represents as much as 58.8% of the lexical bundles that set the French learners apart from at least 5 other learner populations (Section 5.1). Thus, to answer RQ1, over a half of French learners' idiosyncratic use of lexical bundles with verbs can be attributed to L1 influence.

A close look at the lexical bundles and their equivalent forms in French helps identify four major types of transfer effect found in French EFL learners' use of recurrent word sequences, thus addressing RQ2: (1) transfer of collocational and colligational preferences, (2) transfer of syntactic constructions, (3) transfer of functions and discourse conventions and (4) transfer of L1 frequency.

6.1 Transfer of collocational and colligational preferences

In a collocational study of amplifiers in French EFL learner writing, Granger (1998) already interpreted French learners' use of *deeply rooted* as a manifestation of French influence: the collocation has a direct translation equivalent in French, i.e. *profondément enraciné*. Not only is this word combination congruent with *deeply rooted* but the adjective *profondément* is also the most frequent adjective found to modify the past participle *enraciné* in the frWaC (318 occurrences; 0.2 per million). Interestingly, the collocation *firmly rooted* is as frequent as *deeply rooted* in English (0.4 per million in ukWaC) but is not used by French EFL learners. It also has a congruent form in French, i.e. *fermement enraciné*, but this combination is rare (14 occurrences in frWaC; 0.008 per million).

The first language may also prompt learners to use lexical bundles that display untypical colligational patterns in English such as *considered as*. As shown in Examples (3) and (4), French EFL learners mostly use the verb *consider* followed by the preposition *as* to introduce an adjective or a noun phrase (52 occurrences per 100,000 words in ICLE-FR).

(3) Why is this easiness an asset for the EEC to be **considered as** a nation? (ICLE-FR)

(4) Besides childhood is often **considered as** the happiest period in one's life. (ICLE-FR)

French EFL learners' preference for the construction CONSIDER + *as* mirrors the use of French *considérer*, which is typically followed by the preposition *comme* when introducing adjective or noun phrases (Examples (5) and (6)). In frWaC, for example, CONSIDÉRER + *comme* + ADJECTIVE has a relative frequency of 11.5 pmw while the structure without the preposition appears with a relative frequency of 2 pmw.

(5) Il est également **considéré comme** le fondateur de l'abbaye de Malmédy en Belgique. (frWaC)
("He is also considered the founder of the abbey of Malmedy in Belgium.")

(6) La nature a longtemps été **considérée comme** une réserve plutôt que comme un patrimoine. (frWaC)
("Nature has long been considered a reserve rather than a heritage.")

6.2 Transfer of syntactic constructions

Among the lexical bundles that distinguish the French learner population from the other learner groups, several include *to*-infinitive constructions. As illustrated in Example (7),

French learners use the lexical bundle *to go further* although it is not very frequent in English (0.9 pmw in ukWaC). By contrast, the French congruent bundle *aller plus loin* is relatively frequent (8.9 pmw in frWaC).

(7) Nevertheless the Americans decided **to go further** and were the first who wanted to stop Hussein and his army. (ICLE-FR)

The lexical bundle *to be found* appears in several ICLE sub-corpora but it is most frequent in the French learner sub-corpus where it is almost always preceded by a noun phrase (NP) + the verb BE (Examples (8) and (9)). This larger frame corresponds to French NP + ÊTRE + *à trouver*, which is itself a lexical realisation of the frequent French structure NP + ÊTRE + *à* + VERB (over 20 pmw in frWaC). The meaning of this French construction is more commonly expressed with the modal verb *should* in English and the most frequent bundles that exemplify this structure in frWaC include *dossiers sont à retirer* ("forms should be picked up"), *candidatures sont à adresser* ("applications should be sent to"), *précautions sont à prendre* ("precautions should be taken"), *règles sont à respecter* ("rules should be followed"), and *supplément est à payer* ("extra charge should be paid").

(8) The real problem is **to be found** in the fact that women who wish to have a job, also desire to have a family life. (ICLE-FR)

(9) Another example is **to be found** between the French and the Italian vine growers: […]. (ICLE-FR)

There are only two sentences where *to be found* is not used with the verb BE and they both feature the combination *a balance has to be found* (Examples (10) and (11)). Tellingly, the choice of HAVE TO in these two sentences is consistent with the preferred expression of modality in the congruent phrase in French: *un équilibre doit/devra être trouvé* is twice as frequent as *un équilibre est à trouver* in frWaC (41 vs. 23 occurrences).

(10) A balance has thus **to be found**. (ICLE-FR)

(11) And a balance between the two orientations has **to be found**. (ICLE-FR)

Similarly, the lexical bundle *role to play* is always introduced by the verb *have* in ICLE-FR (Example (12)) and this larger word combination is congruent with *avoir un rôle à jouer*, which is the most frequent lexical realisation of the French construction AVOIR + NOUN + *à* + INFINITIVE VERB. This construction is relatively frequent (2.2 per million in frWaC) and lexicalised in a restricted set of recurrent sequences such as *avoir un équilibre à trouver* ("have a balance to find"), *avoir un choix à faire* ("have a choice to make"), *avoir un effort à faire* ("have an effort to make"), *avoir un conseil à donner* ("have advice to give"), and *avoir un défi à relever* ("have a challenge to face").

(12) The parents too have a **role to play** in the education of their children: […] (ICLE-FR)

Transfer of a French lexicalised infinitive construction is also at play in the use of the bundle *is to know whether* in ICLE-FR. As illustrated in Examples (13) and (14), the sequence appears in a larger pattern, i.e. *the question/problem is to know whether*. Strikingly, the French lexicogrammatical pattern *la question/le problème est de* + VERB, is relatively frequent (1.6 pmw in frWaC) and *savoir* is the verb that is most often found in the free slot.

(13) The question **is to know whether** these various agreements will contribute to form a new nation or […] (ICLE-FR)

(14) […] and the problem **is to know whether** reality will be as good as the dream. (ICLE-FR)

6.3 Transfer of functions and discourse conventions

As hypothesized, some of the lexical bundles used idiosyncratically in French EFL learner writing have equivalent forms that fulfil specific discourse functions in French. Quite a few include the first person plural pronoun *we* (*we can say, we can wonder, we may wonder, we must not*) or are part of longer patterns that often involve a personal pronoun subject (*speak of, be tempted to, not forget that, take the example*). First, the ICLE French sub-component is the only L1 learner corpus where the lexical bundle *be tempted to* is found. French EFL learners mostly use the bundle with a modal verb, and with subject pronouns (*we, you*) or the generic noun *people* (cf. Examples (15) and (16)).

(15) After all, we may **be tempted to** believe that this process may at times have been beneficial to the cultural standards that prevail in our society. (ICLE-FR)

(16) Even the most honest people can **be tempted to** satisfy their craving for money. (ICLE-FR)

The larger pattern PRONOUN/GENERIC NOUN (+ MODAL VERB) + *be tempted to* found in ICLE-FR most probably corresponds to two French introductory phrases, i.e. *nous sommes/serions tentés de* and *on est/serait tenté de* (see below for a discussion of EFL learners' use of modal verbs). The pronouns *nous* and *on* are very frequent in French academic writing. The first person plural pronoun *nous* ("we") is commonly used to involve the reader in the argument or guide them through the research process. Such cases of inclusive *we* are often the subjects of procedural verbs (*nous avons procédé à* "we conducted", *nous avons repéré* "we identified") and metadiscursive verbs (e.g. *nous aborderons* "we will discuss", *nous montrerons* "we will show") (Tutin 2010: 38). It may also be found when an argumentative dimension is introduced with an opinion verb (e.g. *penser* "think") or a verb of questioning (e.g. *se demander* "wonder"). With these verbs, however, the indefinite pronoun *on* is much more frequent,[2] especially with the modal verb *pouvoir* (e.g. *on peut admettre* "we can admit", *on peut se demander* "we may wonder") (Tutin 2010: 23).

In the Scientext corpus, the verb *parler* ("speak") is often used in introductory phrases but it is actually found three times as often with the indefinite pronoun *on* as with the personal pronoun subject *nous* ("we") and is modified by *pouvoir* ("can") in 10% of the cases (Example (17)). In the CODIF, by contrast, the two patterns are equally frequent and the more frequent use of *nous* may perhaps be interpreted as a feature of novice writing. When compared to expert writers, for example, French doctoral students have been reported to use more instances of the first plural pronoun subject *nous* in their published research articles (Fløttum & Thue Vold 2010: 46).

(17) Dans ce cas, **on peut parler d**'ellipse métonymique. (Scientext)
("In this case we can speak of metonymic ellipsis.")

These findings help explain French EFL learners' idiosyncratic use of the lexical bundle *speak of* as an effect of their mother tongue. French learners often use the verb with the first person plural pronoun *we* and a modal verb (cf. Example (18)), a pattern that is not common in English academic writing (1.2 pmw in the academic component of the BNC).

(18) **We cannot speak of** a loss of national identity […] (ICLE-FR)

French EFL learners' overuse of lexical bundles including modal verbs is the result of a highly complex interplay of factors. This may, to some extent, simply be a feature of novice writing: both L1 and L2 English student writers are reported to rely extensively on modal verbs to convey statements with an appropriate degree of doubt and certainty (Hyland & Milton 1997). L2 learners, however, appear to depend far more heavily on these devices (e.g. Dagneaux 1995, Granger & Rayson 1998, Aijmer 2002, McKenny 2010) and to have incomplete mastery of the English modal system (Thewissen, 2013).

The difficulties EFL learners face in using modal verbs may be reinforced by interlingual factors as previously reported in the literature for other learner populations. Neff et al. (2003: 216), for example, attribute Spanish and Italian EFL learners' erroneous use of the modal verb *can* in an epistemic sense to a mapping of the more

hypothetical meaning of the Spanish modal verb *poder* and the Italian modal verb *potere* into their L2 English. An unnecessary use of modal verbs may also be associated with transfer of writing conventions from the L1. Neff et al. (2004) explain Spanish learners' overuse of *we must* by the fact that the Spanish modal verb *deber* can mean either *must* or *should* and that *debemos* ("we should" or "we must") + reporting verb is often used as a way of adding a further proposition to be considered by the reader (e.g. *debemos tener en cuenta* "we should/must take into account", *debemos recordar* "we should/must remember", *debemos reconocer* "we should/must recognize", *debemos aceptar* "we should/must accept").

The data analysed for this study contained more examples of transfer of writing conventions. One of the most striking was French EFL learners' overuse of *we can say* (Example (19)), a lexical bundle which is not frequent in English academic writing (0.2 pmw) but is a translational equivalent of both *nous pouvons dire* and *on peut dire* in French (0.3 and 1.3 pmw in Scientext). These two phrases are, among other things, used to introduce the outcome of reasoning or put forward a conclusion in French academic writing (Example (20)) and *on peut dire* is even more frequent in French for general purposes (6 pmw in frWaC).

(19) In conclusion **we can say that** the birth of an economic nation would be favourable. (ICLE-FR)

(20) Dans cette optique, **on peut dire qu'**il existe des genres plus ou moins codifiés …. (Scientext)
("In this perspective, we can say that there are genres which are more or less codified….")

Similarly, the lexical bundle *we may wonder* is absent from the academic component of the British National Corpus and the modal verb *can* is awkward in *we can wonder* (Example (21)). However, both lexical bundles are used by French EFL learners with the meaning and function of the French introductory phrase *on peut se demander* (Example (22)).

(21) But **we can wonder** what a prison is and what its function is in our society. (ICLE-FR)

(22) La question de la compositionnalité sémantique […] et **on peut se demander** si elle présente un intérêt particulier pour le traitement automatique des langues. (Scientext)
("The question of semantic compositionality […] and we can wonder whether it is of particular significance for automatic language processing.")

Learners also use modal verbs of obligation and necessity more than L1 writers and tend to adopt a more direct and emphatic style of persuasion (Hinkel 2002: 110). However, the use of *must*, *should* and *have to* seems to vary widely across different L1 learner populations and reflects at least partly cultural conventions (Hinkel 1995). The lexical bundle *we must not* sets the French learners apart from all the other learner groups except the Swedes. It is used in sequences such as *we must not be pessimistic*, *we must not forget*, *we must not lose sight of*, and *we must not neglect*, to "influence the reader by emotional appeal" (Ädel 2006: 78), persuade them that certain events are desirable, and present the writer and the reader as a team in ICLE-FR (Example (23)).

(23) But **we must not forget that** books used to be written for only a small part of the total population. (ICLE-FR)

(24) Cependant, **il ne faut pas oublier que** les données recueillies auprès des stagiaires sont uniquement déclaratives. (Scientext)

The formally equivalent structure *nous ne devons pas* is not used in French academic writing; neither is the corresponding structure with indefinite *on*, i.e. *on ne doit pas*. To express a negative obligation, French writers rather resort to the impersonal structure *il ne faut pas* (Example (24)) but this discourse strategy is more typical of general rather than academic language (20 pmw in frWaC vs. 1.2 in Scientext). It seems quite probable that French EFL learners' use of the bundle *we must not* is an attempt at

expressing negative obligation and translating *il ne faut pas*, a pattern which is also more frequent in French texts produced by novice writers than expert writers. Interestingly, the larger bundle *we must not forget that* is the only sequence that is repeated in ICLE-FR (5 occ.) and its corresponding structure in French, i.e. *il ne faut pas oublier que*, is also the only lexical bundle that is used repeatedly in Scientext. As illustrated in Example (25), French EFL learners also made use of structures involving the modal verb *should* as functional equivalent patterns to *il ne faut pas oublier*.

(25) **We should not forget that** there are many sorts of criminals, ranging from the accidental criminals and small fry to the hardened ones, the ones "beyond redemption". (ICLE-FR)

The remaining occurrences of the lexical bundle *not forget that* in ICLE-FR are used with the first plural imperative form *let us*, as are a majority of occurrences of the bundle *take the example*. There is no lexically equivalent form to English *let us* in French. Equivalence is however found at the morphological level as French makes use of an inflectional suffix to mark the first imperative plural form. Paquot (2008) compares the use of *let us* in ICLE-FR with that of first person plural imperative verbs in CODIF and finds that the rhetorical and organisational functions fulfilled by *let us* in French EFL learner writing can be paralleled with the very frequent use of first person plural imperative verbs in French student writing to organize discourse and interact with the reader (see also Paquot 2010: 189-191). Imperative forms that are repeated in ICLE-FR often have translational equivalents that are found in CODIF (e.g. *let us take the example of* "prenons l'exemple de"; *let us consider* "considérons"; *let us hope* "espérons"; *let us examine* "examinons"; *let us take* "prenons"; *let us* (*not*/*never*) *forget* "oublions/n'oublions pas que"; *let us think* "pensons"). This generalized overuse of the first person plural imperative in EFL French learner writing as a rhetorical strategy does not conform to English academic writing conventions but rather to French academic style.

Lastly, the use of the lexical bundle *I would say* is also idiosyncratic in ICLE-FR. As shown in Example (26), the bundle is most often used in phraseological 'cascades', "collocational patterns which extend from a node to a collocate and on again

to another node (in other words, chains of shared collocates)" (Gledhill 2000: 212), with an adverbial phrase such as *in conclusion* or *to conclude* to introduce a conclusion.

(26) **In conclusion, I would say that** television has actually replaced religion in our western civilization. (ICLE-FR)

The French bundle *je dirais* appears in Scientext but it is not very frequent (0.38 pmw); the bundle, however, seems to be more typical of informal French and is quite common in frWaC (5.3 pmw). The use of the first person pronoun *je* has long been discouraged in French academic writing but it is used in disciplines such as linguistics (cf. Fløttum 2003, Fløttum et al. 2006), where its use has increased significantly between 1980 and 2000 in research articles (Gjesdal 2003, quoted in Fløttum et al. 2006: 115). The lexical bundle *je dirais* is not found in CODIF but it does occur with a relative frequency of 10 per 10,000 words in the Corpus d'Apprenants du Français Langue Maternelle (CAFLaM), i.e. a newly compiled corpus of argumentative texts produced by French-speaking first year university students (Bolly 2008). Example (27) shows that EFL learners' use of longer sequences and phraseological cascades may also be transfer-related as the bundle *je dirais* is also often introduced by discourse markers such as *en conclusion* ("in conclusion") or *pour conclure* (to conclude") in French-speaking novice writing.

(27) **En conclusion, je dirais qu'**il existe un équilibre à trouver entre conformisme et différence. (CAFLaM)
("In conclusion, I would say that a balance should be found between conformism and difference.")

It may thus be argued that French EFL learners' use of *I would say* is the result of a combination of L1-related factors, i.e. the relative and increasing tolerance of *je* in the discipline they are studying, the high frequency of *je dirais* in general language, and French-speaking novice writers' reliance on phraseological cascades including *je dirais* to conclude their argumentative essays.

6.4 Transfer of L1 frequency

Congruency is not a sufficient factor for cross-linguistic influence and distributional properties in the first language seem to play a significant role as well. Another way of approaching L1 frequency is to check whether the lexical bundles used by the French learners have equivalent structures in another language represented in the ICLE corpus and if so, why these are not transferred into English by the other learner population. Spanish is arguably a good candidate for this purpose: French and Spanish are both Romance languages, and there are often congruent sequences in Spanish for the French word combinations that were pinpointed in this study as responsible for transfer effects.

Spot-checks in the 100 million word Web corpus of Spanish available in the Sketch Engine indeed strengthen the case for transfer of L1 frequency. While CONSIDERAR + *como* ("CONSIDER as") exists in Spanish, for example, the verb is much more frequently used without the preposition and the pattern CONSIDERAR + ADJECTIVE is ten times as frequent as CONSIDERAR *como* + ADJECTIVE (49.8 vs. 4.7 pmw). As a result, Spanish learners sometimes use the preposition *as* after the verb *consider* (10 occurrences per 100,000 words in ICLE-SP) but they are much less tempted to do so than their French counterparts. Similarly, *el problema es de* + VERB ("the problem is to + VERB") is extremely rare (0.008 pmw) in the Spanish web-derived corpus and only two instances of *the problem/question is to* + VERB are found in the Spanish learner corpus (Examples (28) and (29)).

(28) The second and greatest **problem is to** perform what they have learnt. (ICLE-SP)

(29) Another important **question is to** analyze what can the goverment do about this problem because they must fight to find a solution. (ICLE-SP)

A comparison between French and Spanish also supports an L1 frequency-based explanation for the overuse of the lexical bundle *role to play* in ICLE-FR and its absence in ICLE-SP. There is a congruent form in Spanish (*papel que desempeñar*) but

it is rare (0.12 pmw), as is the larger construction TENER + NP + *que* + INFINITIVE VERB ("HAVE + NP + to + INFINITIVE VERB") (0.4 pmw).

Spanish academic writing is characterized by a *we*-stance and as a result, EFL Spanish learners also tend to overuse introductory phrases with *we can* (Neff et al. 2001). Patterns of overuse in ICLE-SP are however less marked when compared to ICLE-FR. A likely explanation for this lies in the fact that first person plural indicative forms compete with *se* impersonal passive phrases to perform similar discourse functions in Spanish. If we look at the Spanish translational equivalents of the English lexical bundles that are characterised by transfer of discourse conventions in French learner writing, the prominent role of *se* impersonal passive structures appears clearly: *podemos preguntarnos* ("we may wonder") and *se puede preguntar* ("it can be wondered") are equally frequent (0.4 and 0.3 pmw); *se puede decir* ("it can be said") is slightly more frequent than *podemos decir* ("we can say") (11.1 vs. 10.5) and *se puede hablar de* appears 251 times while *podemos hablar de* ("we can speak of") only occurs twice in the corpus. The reason why Spanish learners use fewer *we can* constructions than French learners is therefore most probably because they have the choice between this construction and impersonal phrases with *se*. This mirrors Neff van Aertselaer's (2008) claim that Spanish expert and novice writers' use of passive structures in English probably reflects a transfer from Spanish discourse strategies.[3]

To sum up, congruency or formal equivalence is often pinpointed as the explanatory factor for transfer effects. The results presented here show that congruency is not sufficient in itself to ensure that a formal equivalent word combination will be used in the foreign language. It is not because there is a formal equivalent of an English lexical bundle in French and Spanish that the two learner populations will use the English bundle in the same way. The frequency of word combinations in the first language seems to play a crucial role: the more frequent a lexical bundle is in the learners' mother tongue, the more likely learners are to use its congruent form in the foreign language. This seems to hold true for lexical bundles that exemplify collocations (*deeply rooted*), colligations (*consider as*), syntactic structures (NP + *to*-infinitive, e.g. *role to play*; NP + *is* + *to*-infinitive, e.g. *the question is to know whether*) and discourse conventions (*we*-lexical bundles) alike.

6. Conclusion

Transfer effects on French learners' use of 3-word sequences with lexical verbs do not seem to generate obvious errors, at least at the intermediate to advanced proficiency levels represented in the French component of the International Corpus of Learner English. Rather, they are more visible in the learners' selection of unmarked word combinations whose translational equivalents are deeply entrenched in French speakers' mental lexicon because these sequences are particularly frequent or are directly anchored to important communicative or metatextual functions. The word strings may be typical English sequences (e.g. *deeply rooted*) or less favoured combinations (e.g. *considered as*). More interestingly perhaps, they may be perfectly correct combinations in English but more commonly used in less formal genres than that of academic writing: in the British National Corpus, the lexical bundles *I would say that*, *we can say*, *we must not*, *let us not forget that*, and *let us take the example* are generally more frequent in non-academic texts and speech varieties including lectures and meetings.

All in all, results are in line with a usage-based view of language that recognizes the active role that the L1 may play in the acquisition of a foreign language (e.g. Bybee 2008). EFL learners bring knowledge of the L1 lexicon to the writing task in the foreign language, including preferred collocations and lexicogrammatical patterns of words, as well as their stylistic or register specificities, discourse functions and frequency of use. As put by Hoey (2005),

> As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. The same applies to word sequences built out of these words; these too become loaded with the contexts and co-texts in which they occur.
>
> (Hoey 2005: 8)

The transfer effects identified in this study are thus best described as "transfer of primings" (Hoey 2005: 183). Mental primings for (at least frequent or core) L1 words

and word strings are most probably superimposed on the primings for their translation equivalent forms in the foreign language.

The direct pedagogical implication is that EFL teaching needs to counter the default and sometimes misleading L1-related primings in EFL learners' mental lexicons. Awareness-raising activities focusing on similarities and differences between the mother tongue and the foreign language are clearly needed. They should not be restricted to "helping learners focus on errors typically committed by learners from a particular L1" (Hegelheimer & Fisher 2006: 259) but should also raise learners' awareness of more subtle differences such as the collocational preferences and distributional properties of similar words in the two languages. This recommendation stands in sharp contrast to Bahns's (1993: 56) claim that collocations which are direct translation equivalents do not need to be taught. Learners have no way of knowing which collocations are congruent in the mother tongue and the foreign language; moreover, the differences between the collocations in L1 and L2 may lie in aspects of use rather than form or meaning.

Primings are also sensitive to the textual, generic and social contexts in which a lexical item is encountered. Hoey (2005: 10) illustrates this with the word *research*, which is primed in the mind of academic language users to occur with *recent* in academic discourse and news reports of research but is not primed to occur in other text types or other contexts. A direct implication of Hoey's theory of lexical priming is that academic-like word combinations in the first language cannot be assumed to be primed in the mental lexicon of novice native writers who may have had little contact with academic texts in their L1. While many of the French lexical bundles examined here proved to be relatively frequent in French academic writing, some of them are indeed primed more strongly in general language. This is particularly true of two sequences, i.e. *on peut dire* and *il ne faut pas*, and calls for a more systematic deconstruction of the concept of L1 frequency in future research.

Many learner corpus-based studies, however, have fallen into the trap of claiming L1 influence on the basis that the structure exists in the first language without further investigation of L1 empirical data. In Douglas's (2001: 451) words, "the point here is not that these methods are faulty or that the interpretations are invalid, but only that little or no evidence is provided for either quality [reliability and validity]". As

shown in this study, formal similarity between L1 and L2 word combinations does not necessarily make the word combination in the first language a strong candidate for transfer into the foreign language. Other factors intervene and L1 frequency proved to contribute to transferability in a significant way. The impact of L1 frequency is most apparent when different languages are compared with the help of corpus data. As a consequence, this study also brings support to the detection-based approach to transfer first outlined in Jarvis (2010). The method is based on the premises that it is possible to identify the first language of a learner on the basis of their use of specific features of the target language and that these idiosyncrasies can serve as useful indicators of cross-linguistic influence (Jarvis 2012).

Transfer effects were indeed pinpointed for twenty 3-word lexical bundles which were further analysed as part of fifteen longer strings. This represents 7.3% of all the 3-word sequences that appear at least 5 times in the French learner corpus and c. 60% of the bundles that set the French learners apart from at least 5 other learner populations. These figures are already quite high but they certainly underestimate the impact of the first language. The criterion according to which French learners' use of a given lexical bundle has to differ from that of at least five other learner groups is very conservative. L1 influence may be obscured when the effects of the mother tongue of different L1 learner populations coincide to produce the same IL behaviour and this is certainly not a rare phenomenon (Jarvis 2000).

More generally, the study has also brought to light the considerable potential of a corpus-driven approach to track L1 influence on learner language. Transfer studies have often investigated "bits and pieces of learners' language chosen for analysis because they caught the researcher's eye, seemed to exhibit some systematicity, confirmed some intuition one had about SLA, or had been found interesting in L1 acquisition" (Lightbown 1984: 245). As put by De Cock (2004), the lexical bundle approach represents "corpus linguistic methodology at its most heuristic, i.e., as a raw discovery procedure" (De Cock 2004: 227). Coupled with Jarvis's (2000) framework and appropriate statistical tests, it proves most useful to extract fully automatically a number of word combinations that deserved further analysis and consequently identify transfer effects that until now have been little documented in the SLA literature. Lexical transfer has too often been narrowed down to transfer of form/meaning mappings and

the third aspect of word knowledge, i.e. use, has rarely been investigated in all its complexity. Further research is clearly needed. Lexical bundles of different sizes and built around different word classes than just verbs should prove fascinating data types to start with.

Notes

***** I would like to thank Sylviane Granger, Victoria Hasko and two anonymous reviewers for their valuable comments and constructive suggestions for improvement. I acknowledge the financial support of the Fonds de la Recherche Scientifique (FNRS).

1. The use of parametric tests may be criticized as the data used in this study is not normally distributed. According to Howell (1997), those who argue in favour of using parametric tests "argue, however, that the assumptions normally cited as being required of parametric tests are overly restrictive in practice and that the parametric tests are remarkably unaffected by violations of distribution assumptions" (Howell 1997: 646, see also Rietveld et al. 2004: 360). Moreover, parametric tests are said to be more powerful than non-parametric tests: they require fewer observations than do non-parametric tests and are more likely "to lead to rejection of a false null hypothesis" (Howell 1997: 646) than are their corresponding non-parametric tests. This advantage seems to be maintained "even when the distribution assumptions are violated to a moderate degree" (ibid).

2. The French indefinite pronoun *on* is much more frequent and stylistically very different from the English *one*: it can refer to one or more people, be substituted for all personal pronouns and "has an unclear enunciative status (i.e. relation to speaker or locator and receiver)" (Fløttum et al. 2006: 113).

3**.** Paquot (2008) has also shown that the distribution of *let us* in the interlanguage of French, Spanish and Dutch learners parallels that of first person plural imperative structures in the three languages.

References

Ädel, A. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.

Ädel, A. 2008. "Involvement features in writing: Do time and interaction trump register awareness". In B. M. Diez-Bedmar, G. Gilquin & S. Papp (Eds.), *Linking up Contrastive and Learner Corpus Research*. Amsterdam/New York: Rodopi, 35–53.

Aijmer, K. 2002. "Modality in advanced Swedish learners' written interlanguage". In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 55–76.

Allen, D. 2011. "Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus". *Komaba Journal of English Education*, 1, 105–127

Altenberg, B. 1998. "On the phraseology of spoken English: The evidence of recurrent word-combinations". In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 101–122.

Bahns, J. 1993. "Lexical collocations: A contrastive view". *ELT Journal*, 47 (1), 56–63.

Biber, D., Conrad, S. & Cortes, V. 2003. "Lexical bundles in speech and writing: An initial taxonomy". In A. Wilson, P. Rayson & T. McEnery (Eds.), *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech*. Frankfurt: Peter Lang, 71–92.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bolly, C. 2008. *Corpus Argumentatif en Français Langue Maternelle (CAFLaM). Les Unités Phraséologiques: Un Phénomène Linguistique Complexe?*, Unpublished Ph.D Thesis, Louvain-la-Neuve: Université catholique de Louvain.

Bybee, J. 2008. "Usage-based grammar and second language acquisition". In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge, 216–236.

Cortes, V. 2004. "Lexical bundles in published and student disciplinary writing: Examples from history and biology". *English for Specific Purposes*, 23 (4), 397–423.

Dagneaux, E. 1995. *Expressions of Epistemic Modality in Native and Non-Native Essay-Writing*. MA dissertation. Louvain-la-Neuve: Université catholique de Louvain.

De Cock, S. 2003. *Recurrent Sequences of Words in Native Speaker and Advanced Learner Spoken and Written English: A Corpus-Driven Approach*. Unpublished PhD thesis. Louvain-la-Neuve: Université catholique de Louvain.

De Cock, S. 2004. "Preferred sequences of words in NS and NNS speech". *Belgian Journal of English Language and Literatures (BELL), New Series*, 2, 225–246.

Douglas, D. 2001. “Performance consistency in second language acquisition and language testing”. *Second Language Research*, 17 (4), 442–456.

Fløttum, K. 2003. “Personal English, indefinite French and plural Norwegian scientific authors? Pronominal author manifestation in research articles”. *Norsk Lingvistisk Tidsskrift*, 21 (1), 21–55.

Fløttum, K., Dahl, T. & Kinn, T. 2006. *Academic Voices – Across Languages and Disciplines.* Amsterdam: John Benjamins.

Fløttum, K. & Thue Vold, E. 2010. “L’éthos auto-attribué d’auteurs-doctorants dans le discourse scientifique”. *Lidil, Revue de Linguistique et de Didactique des Langues*, 41, 41–58. Available at: http://lidil.revues.org/index3006.html (accessed June 2013).

Gjesdal, A. M. 2003. *L’emploi du Pronom ‘on’ dans les Articles de Recherche. Une étude diachronique et qualitative*. MA dissertation. Bergen: University of Bergen.

Gledhill, C. 2000. *Collocations in Science Writing*. Tuebingen: Gunter Narr Verlag.

Granger, S. 1998. “Prefabricated patterns in advanced EFL writing: Collocations and formulae”. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications.* Oxford: Oxford University Press, 145–160.

Granger S., Dagneaux E., & Meunier F. 2002. *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. 2009. *The International Corpus of Learner English. Handbook and CD-ROM* (Version 2). Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S. & Rayson, P. 1998. “Automatic profiling of learner texts”. In S. Granger (Ed.), *Learner English on Computer*. London/New York: Addison Wesley Longman, 119–131.

Groom, N. 2009. “Effects of second language immersion on second language collocational development”. In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language*. Basingstoke: Palgrave Macmillan, 21–33.

Hegelheimer, V. & Fisher, D. 2006. “Grammar, writing, and technology: A sample technology-supported approach to teaching grammar and improving writing for ESL learners”. *CALICO Journal*, 23 (2), 257–279.

Hinkel, E. 1995. “The use of modal verbs as a reflection of cultural values”. *TESOL Quarterly*, 29 (2), 325–243.

Hinkel, E. 2002. *Second Language Writers’ Text: Linguistic and Rhetorical Features*. London: Lawrence Erlbaum Associates.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London/New York: Routledge.

Howell, D. 1997. *Statistical Methods for Psychology*. Belmont: Wadsworth.

Hyland, K. & Milton, J. 1997. "Qualifications and certainty in L1 and L2 students' writing". *Journal of Second Language Writing*, 6 (2), 183–205.

Jarvis, S. 2000. "Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon". *Language Learning*, 50 (2), 245–309.

Jarvis, S. 2010. "Comparison-based and detection-based approaches to transfer research". In L. Roberts, M. Howard, M. Ó Laoire & D. Singleton (Eds.), *EUROSLA Yearbook 10*. Amsterdam: John Benjamins, 169–192.

Jarvis, S. 2012. "The detection-based approach: An overview". In S. Jarvis & S. Crossley (Eds.), *Approaching Language Transfer through Text Classification: Explorations in the Detection-Based Approach*. Bristol: Multilingual Matters, 1–33.

Juknevičienė, R. 2009. "Lexical bundles in learner language: Lithuanian learners vs. native speakers". *Kalbotyra* 61 (3), 61–72.

Kellerman, E. 1978. "Giving learners a break: Native language intuitions as a source of predictions about transferability". *Working Papers on Bilingualism*, 15, 59–92.

Kilgarriff, A. & Kosem, I. 2012. "Corpus tools for lexicographers". In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford University Press, 31–56.

Kilgarriff, A., Rychly, P., Smrz P. & Tugwell, D. 2004. *The Sketch Engine. Proc EURALEX*. Lorient, France, 105–116. Available at: www.sketchengine.co.uk (accessed June 2013).

Kroll, B. 1990. "What does time buy? ESL student performance on home vs. class compositions". In B. Kroll (Ed.), *Second Language Writing*. Cambridge: Cambridge University Press, 140–154.

Lightbown, P. M. 1984. "The relationship between theory and method in second-language-acquisition research". In A. Davies, C. Criper & A. Howatt (Eds.), *Interlanguage*. Edinburgh: Edinburgh University Press, 241–252.

McKenny, J. 2010. *A Corpus Study of the Phraseology of Written Argumentative English*. Saarbrücken: Lambert Academic Publishing.

Neff, J., Ballesteros, F., Dafouz, E., Martínez, F. & Rica, J. P. 2004. "The expression of writer stance in native and non-native argumentative texts". In R. Facchinetti & F. Palmer (Eds.), *English Modality in Perspective*. Frankfurt am Main: Peter Lang, 141–161.

Neff, J., Dafouz, E., Herrera, H., Martínez, F., Rica, J. P., Diez, M., Prieto, R. & Sancho, C. 2003. "Contrasting learner corpora: The use of modal and reporting verbs in the expression of writer stance". In S. Granger & S. Petch-Tyson (Eds.), *Extending the Scope of Corpus-Based Research. New Applications, New Challenges*. Amsterdam/New York: Rodopi, 211–230.

Neff, J., Martínez, F. & Rica, J. P. 2001. "A contrastive study of qualification devices in NS and NNS argumentative texts in English". In ERIC Clearing House on Language and Linguistics (ERIC Document Reproduction Service, ED 465301). Washington, D.C.: Educational Resource Information Center, U.S. Department of Education.

Neff van Aertselaer, J. 2008. "Contrasting English-Spanish interpersonal discourse phrases: A corpus study". In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins, 85–100.

Paquot, M. 2008. "Exemplification in learner writing: A cross-linguistic perspective". In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: Benjamins, 101–119.

Paquot, M. 2010. *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. London/New York: Continuum.

Paquot, M. forthcoming. "Phraseology and lexicography". In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of Corpus Linguistics*. Cambridge: Cambridge University Press.

Paquot, M. & Granger, S. 2012. "Formulaic Language in Learner Corpora". *Annual Review of Applied Linguistics*, 32, 130–149.

R Core Team. 2012: online. *R: A Language and Environment for Statistical Computing*. Available at: http://www.R-project.org (accessed June 2013).

Rayson, P. 2003. *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Unpublished PhD thesis, Lancaster University.

Reppen, R. 2009. "Exploring L1 and L2 writing development through collocations: A corpus-based look". In A. Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language*. Basingstoke: Palgrave Macmillan, 49–59.

Rica, J.P. 2010. "Corpus analysis and phraseology: Transfer of multi-word units". *Linguistics and the Human Sciences*, 6, 321–343.

Rietveld, T., van Hout, R. & Ernestus, M. 2004. "Pitfalls in corpus research". *Computers and the Humanities*, 38 (4), 343–362.

Ringbom, H. 1987. *The Role of the First Language in Foreign Language Learning*. Clevedon/Philadelphia: Multilingual Matters.

Scott, M. 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

Selinker L. 1992. *Rediscovering Interlanguage*. London/New York: Longman.

Scientext corpus. Available at: http://scientext.msh-alpes.fr/ (accessed June 2013)

Thewissen, J. 2013. "Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus". *Modern Language Journal*, 97 (Suppl. 1), 77–101.

Tutin, A. 2010. “Dans cet article, nous souhaitons montrer que… Lexique verbal et positionnement de l’auteur dans les articles en sciences humaines. Enonciation et rhétorique dans l’écrit scientifique”. *Lidil, Revue de Linguistique et de didactique des langues*, 41, 15–40. Available at: http://lidil.revues.org/index3040.html (accessed June 2013)

Author's address

Magali Paquot
Centre for English Corpus Linguistics
Institut Langage et Communication
Université catholique de Louvain
Place Blaise Pascal 1, bte L3.03.31
1348, Louvain-la-Neuve
Belgium

magali.paquot@uclouvain.be