High-frequency verbs: starting block or stumbling block for advanced L2 communication? Insights from native and learner corpora

Claire Hugon (F.R.S.-FNRS)

Draft version

1. Introduction

High-frequency verbs are tricky items for language learners. They are essential to basic communication, yet they represent such a heavy learning burden that, even after years of learning, some of their more complex uses remain difficult to grasp. This article investigates how advanced learners cope with these verbs, by means of an in-depth corpus-based analysis of *MAKE* in native and advanced French-speaking learner speech and writing. First, advanced interlanguage will be discussed: this lesser-known variety of learner language has in recent years enjoyed a regain of attention thanks to the emergence of computerized learner corpora, but we will see that it is still somewhat hazy. The next section will describe high-frequency verbs and what it is that makes them such difficult items for learners. This will be followed by a case study of *MAKE,* which will examine both qualitative and quantitative aspects of the verb. The article will be rounded off by a conclusion including pedagogical implications and suggestions for a more adequate treatment of high-frequency verbs in foreign language teaching.

2. The haziness of advanced interlanguage

For decades, researchers working in the field of Second Language Acquisition (SLA) viewed learner language as nothing more than a defective variant of native language. The deviance from the native norm was seen as the result of errors, largely if not wholly determined by the learner's mother tongue (L1). The extent of this deviance could allegedly be predicted by examining similarities and differences between L1 and the target language (L2). In the late 1960s and early 1970s, however, the wind

turned: in his seminal 1967 paper "The Significance of Learners' errors", Corder underlined the importance of distinguishing "systematic and non-systematic errors" and argued that the former reveal the learner's "transitional competence", that is, his "underlying knowledge of the language to date" (Corder 1967:166). They are, in James's words (1998:78), "not self-corrigible" by the learner. Errors of performance, or "mistakes", are "of no significance to the process of language learning" (Corder 1967:167); they are just slips that the student is able to correct if given the opportunity. A few years later Selinker (1972), who had become very much concerned with examining learner language in its own right, coined the word 'Interlanguage' (IL) to refer to learner language as a system which is neither the L1, nor the L2, but something in between and constantly evolving with the learner: a language with its own rules, its own grammar and its own lexicon, worthy of being described sui generis. IL, instead of L1, became the starting point for analysis (de Bot et al. 2005:34). Each learner has his or her own interlanguage, but learners tend to build it up in a largely predictable way, using a number of learning strategies to move through successive steps on the IL scale.

Early L2 acquisition was, from the early 1970s onwards, the object of a flurry of studies devoted, most notably, to the acquisition order of morphemes (e.g. Dulay & Burt 1973, 1974) but also to the acquisition sequence of items such as question formation (e.g. Huang & Hatch 1978), negative structures (e.g. Milon 1974) and relative clauses (e.g. Gass 1979).

Research on advanced L2 acquisition, on the other hand, is much scarcer (de Haan 1998, Cobb 2003, Granger 2004). Cobb (2003:394-395) suggests two main reasons why "intermediate-advanced IL remains relatively unchartered". The first reason is that, until recently, researchers simply lacked the data they needed to conduct such investigations. Not only are intermediate and advanced learners less easy to observe than, for example, school children beginners, but there was no practical way of collecting and analyzing learner production. While this is no longer an issue, thanks to the growing availability of learner corpora, another problem remains: late interlanguage, no doubt due to the aforementioned shortage of data, has not been described adequately in the SLA literature. However, late IL has recently been the object of a number of studies: contrary to SLA specialists, corpus linguists interested in learner language have focused primarily on precisely those varieties of interlanguage that lack a strong theory, but that are well represented in learner

corpora. This means that what information we have about advanced learner language comes from corpus rather than SLA circles (Granger 2004:134). As a result, the studies in question have tended to confine themselves to the descriptive level, with very little interpretation of the data and therefore little contribution to the fields of foreign language learning and teaching. In Granger's (2004:135) words, this probably stems from the fact that corpus linguists "felt the need to establish the facts first before launching into theoretical generalisations".

An additional difficulty in defining advanced learner language is that, at this stage, patterns of acquisition are not as easy to pin down as at lower proficiency levels (see e.g. Bartning 1997). While syntactic and morphological errors are easy enough to identify, linguistic features typical of advanced interlanguage are often less overt and typically involve "usage, lexical choice, stylistic appropriacy, and various sorts of global discourse errors"; in other words, features "which are not necessarily completely erroneous" (Lennon 1991:183-184).

Computer learner corpora have proved a great help in this context: in addition to errors, they reveal over- and underuse of L2 items, two tendencies described by Bartning (1997:42-44) as characteristic of advanced interlanguage. As Granger (2002:4) puts it, we "have very little intuitive awareness" of features such as frequency, yet over- and underuse also contribute to the non-native-like character of learner production. Moreover, consisting as they do of whole texts (instead of being "depositories of errors", Granger 1998:6), computer learner corpora do not focus only on what learners get wrong: they also enable us to investigate non-native language from a positive point of view and to see what learners get right (Leech 1998:xvii).

Despite previous mutual wariness, there are encouraging signs that SLA specialists and learner corpus linguists are gradually acknowledging each other's value (Granger in press), as they realize the advantages which can come of bringing together their respective assets. Well-designed learner corpora now take SLA findings into account, notably by recording as many variables as possible, and linking to each text a detailed profile of the speaker/writer. In this way, the social, individual and cognitive factors which have been shown to play a role in second language acquisition (see e.g. Ellis 1994) are an integral part of any well-designed learner corpus. Computerized corpora, for their part, can contribute to SLA research by giving access to an unprecedented amount of naturally occurring data, thus bringing a welcome answer to the problematic issues of representativeness and generalizability which

arose from the restricted empirical basis of SLA studies so far. As we accumulate descriptions of advanced IL (McLaughlin 1987:80) and combine the strengths of SLA and computer learner corpora, we can hope to gain a better understanding of the nature of advanced learner language. In turn, if we know what difficulties advanced learners still encounter, and understand the acquisition sequences they are moving through, then "instruction can be focused more effectively throughout the learning process" (Cobb 2003:396) and we can tailor teaching to learner needs. Foreign Language Teaching (FLT) is, after all, one of the most obvious applications of SLA, "since SLA researchers study the process language teaching is designed to facilitate" (Doughty & Long 2003:7). After this first general overview of late interlanguage, we will now focus on an area of language that is notoriously still perilous for proficient learners, i.e. high-frequency verbs.

3. High-frequency verbs

As has just been mentioned, high-frequency verbs have been shown to remain problematic for learners, even at the advanced level (e.g. Källkvist 1998, Altenberg & Granger 2001, Nesselhauf 2005). At first glance, one could wonder why such core items should still be difficult after the first stages of acquisition: after all, they are learnt very early in the curriculum, as they are essential to communication. Yet this might be part of the reason why learners still struggle with them after years of teaching: these verbs are then neglected on the assumption that they are known. Consequently, learners tend to have a truncated knowledge of these verbs and to rely essentially on their basic uses. These uses are then hugely overused, as learners (1) clutch for items they feel safe with (what Hasselgren (1994:237) memorably termed the "teddy-bear effect") and (2) make do with fewer words and set up high-frequency verbs as hyperonyms (a strategy highlighted by Levenston & Blum 1976). These two strategies of communication, working together and reinforcing one another, lead to a massive overuse of high-frequency verbs (Hasselgren 1994; Granger 1996; Källkvist 1999).

Yet this overuse, induced by an ill-advised sense of security, only increases the already numerous opportunities for learners to use these verbs erroneously. Under their deceptively safe appearance, high-frequency verbs are treacherous for learners,

even at the advanced level. Like other very frequent words, they are fraught with difficulties: not only are they likely to be only partially learned because of the multiplicity of features to be acquired (Laufer 1997:142), but the context of use causes them to "shift in meaning in subtle and unpredictable ways" (Hoey 2002). As high-frequency verbs are not targeted much in later stages of teaching, learners' knowledge of their more complex and idiomatic aspects leaves much to be desired. This probably explains Sinclair's claim (1991:79) that learners tend to avoid very frequent verbs "especially when they make up idiomatic phrases. Instead of using them, they rely on larger, rarer, and clumsier words which make their language sound stilted and awkward". Overuse and underuse of these verbs are not contradictory observations (Altenberg & Granger 2001:174), they concern different aspects of the words.

Nation (1990:30-49) discusses what is involved in knowing a word and argues that productive knowledge of a word involves, among other things, knowing its pronunciation and spelling, its grammatical patterns and usual collocations, as well as its frequency and contexts of use. Even native speakers, he remarks, probably only achieve this level of knowledge in a small proportion of their lexicon. The variables that influence the learning load are, according to Nation, (1) the learners' L1 and their previous experience of the L2, (2) teaching (which can have a positive, neutral, or negative effect) and (3) intrinsic difficulty. Laufer (1997) discusses in more detail a number of intralexical factors that affect vocabulary acquisition. These include, among others, morphology, synformy, part of speech, and semantic features. The latter are further subdivided into abstractness, specificity and register restrictions, idiomaticity and multiplicity of meaning. To sum up, the amount of effort required to learn a lexical item is a function of (1) intrinsic properties of the item, (2) the learners' L1, (3) the learners' proficiency level in the L2, and (4) teaching.

As regards intralexical variables, high-frequency verbs are particularly complex from a syntactic and semantic point of view. First, they illustrate perfectly Van Roey's (1990:56) remark that "in a general way, the more frequent a word is, the more polysemous it tends to be". This polysemy, according to Altenberg & Granger (2001:174), is caused by two kinds of meaning extension – both no doubt resulting from the extremely high frequency of occurrence of these verbs: "one universal tendency creating more general, abstract, delexicalized or grammaticalized uses" - in other words, depletion of meaning – and "various language-specific tendencies

resulting in specialized meanings, collocations, and idiomatic uses" – or contextualisation of meaning. These two tendencies result in numerous phraseological uses, where constraints of use are more or less arbitrary. In many instances, semantics and phraseology are inseparable: the meaning of the verb is inextricably linked to the specific context in which it is used. To complicate matters, high-frequency verbs also enter into an uncommonly large number of phrasal verbs built with various prepositional and adverbial particles.

To top it all, high-frequency verbs are also extremely register-sensitive: they are, for example, much more frequent in conversation than in fiction, newspapers or academic writing (Biber et al. 1999:373). Moreover, research has shown that the various patterns, meanings and functions of these verbs are distributed unevenly across genres. The verbs *MEAN, KNOW, THINK,* and *SEE*, for instance, have been found to appear very frequently in informal speech, often in typically spoken structures that are unlikely to ever crop up in formal writing (e.g. *you know* and *I mean* used as discourse markers, see Biber et al. 1999:359-360).

These intralexical difficulties hold true for foreign language learners, who are faced with the additional difficulty of translational equivalence. In Lennon's words (1996:25), the already considerable breadth of knowledge that learners must develop "may be contaminated by the status of the word's perceived 'translational equivalent' in L1 or other known languages". High-frequency verbs typically have a major translational equivalent in most languages, but the cross-linguistic similarity is mostly deceptive, as shown among others in studies by Viberg (1996) and Altenberg (2001). In a study of the cognates English *go* and Swedish *gå* Viberg (1996) demonstrates that these cognates only correspond to one another in about one third of the occurrences he examined. Thus, although they are "generally regarded as translational equivalents, they are in fact rendered by other verbs in the majority of cases in translations between the languages" (Altenberg & Granger 2001:193). Viberg (1996:161) remarks that "a relatively low proportion of congruent cases is probably characteristic of nuclear verbs and other highly exploited words". Yet another problematic aspect of the diverging polysemy of high-frequency verbs is the arbitrariness of phraseological uses: the choice of the verb is seldom semantically motivated, and does not necessarily correspond cross-linguistically. This diverging polysemy can thus be treacherous for foreign language learners. They are likely to overgeneralise the

equivalence (because they erroneously assume the use of the L2 verb to be identical to that of its L1 correspondent), which will lead to non-native-like uses.

Clearly, "knowing a word" is not an all-or-nothing matter, and it is unwise to consider that basic verbs have been appropriately dealt with when students have been taught their basic meanings. This often results in learners having "only a very crude knowledge of their grammatical and lexical patterning" (Altenberg & Granger 2001:190). Gilquin (2007:287) confirms, for example, that learners' knowledge of high-frequency verbs "tends to be rather superficial, learners being essentially familiar with the core meanings of the verbs, but largely ignorant of their collocates".

4. Case study: *MAKE* in native and learner speech and writing

In this section, we will zoom in on one high-frequency verb, viz. the verb *MAKE*. Its use will be analysed in the writing and speech of native and French-speaking students. *MAKE* belongs to the most frequent verbs in native English. It is also characterized by a high degree of polysemy: in their detailed study of the treatment of *MAKE* in five monolingual learner's dictionaries, De Cock and Granger (2004) show that the number of independent meanings ranges from six to fourteen according to the dictionary. *MAKE* has near-equivalents in many languages, including French *FAIRE*. It is worth mentioning, however, that *FAIRE* and *MAKE* display diverging polysemy, the most obvious illustration of this being that *FAIRE* can be rendered in English by both *MAKE* and *DO* depending on the context. This is liable to result in confusion in learner production, as is indeed reported by Borgatti (2006:92): according to his results, mix-ups between *MAKE* and *DO* are responsible for 14.7% of erroneous instances of *MAKE* and 62.5% of erroneous occurrences of *DO*. *MAKE* also enters into a wide range of restricted and idiomatic uses: Howarth (1996:119) reports an extremely high rate of restricted collocations and idioms for *MAKE* in the V + N pattern in native academic writing: 352 out of the total 369 occurrences display some degree of restrictedness, which corresponds to more than 95%. Furthermore, *MAKE* is the basis for a large number of phrasal and prepositional verbs, extending its semantic scope even further: the *Macmillan English Dictionary for Advanced Learners* (*MEDAL*, 2007) lists no less than thirteen different phrasal/prepositional verbs built

with *MAKE* (given that some of these are themselves polysemous, the number of additional meanings reaches 37 in the *MEDAL*).

4.1. Data and methodology

The present study relies on four corpora compiled by the Centre for English Corpus Linguistics (CECL) at Louvain University, as shown in Table 1: the French-speaking component of the International Corpus of Learner English (Granger et al. 2002) – or ICLE-FR[1] - and its native counterpart, the Louvain Corpus of Native English Essays (LOCNESS) provided the written data, while the spoken data came from the French subcorpus of the Louvain International Database of Spoken English Interlanguage (LINDSEI-FR) and its control corpus, the Louvain Corpus of Native English Conversation (LOCNEC). These corpora are fully comparable in terms of participants (university undergraduate students of approximately the same age) and consist of argumentative essays (for ICLE-FR and LOCNESS) and informal interviews (for LINDSEI-FR and LOCNEC)[2].

Corpus	Number of running words
Native speakers (NS)	
LOCNESS	150,590
LOCNEC	118,555
Non-native speakers (NNS)	
ICLE-FR	160,784
LINDSEI-FR	90,851

Table 1: Size of the four corpora

All the inflectional forms of *MAKE* (*make/makes/making/made*) were retrieved from the four corpora using the *Concord* tool of the *Wordsmith Tools 4* package (Scott 1999). This generated a concordance of the lemma *MAKE* for each corpus. All irrelevant examples were then weeded out manually by scrutinizing each concordance line. In order to make the overview as comprehensive as possible, occurrences such as *making* used as a noun (1), compound uses (2) or phrasal verbs (3) were included.

---

[1] Only the argumentative essays were retained here (ICLE also contains a number of literary essays).
[2] For the spoken corpora, only the interviewee's words (the 'b-turns') were kept, that is, the interviewers' turns (the 'a-turns') were deleted.

(1) The **making** of Europe on a social level will... (ICLE-FR)

(2) ... racism in our Universities is not a **make-believe** problem. (LOCNESS)

(3) How about **making** the paycheck **out** to the couple... (LOCNESS)

On the other hand, there were several occurrences of the noun *make-up* (4), which were excluded because its relationship with *MAKE* was felt to be too tenuous.

(4) ... it's like you don't notice whether their **make-up**'s good or not... (LOCNEC)

The case study is structured in two main stages: first, the overall results for the four corpora will be presented and discussed, after which a more qualitative approach will be presented, which will investigate the polysemy of *MAKE* by looking at how its main meanings and uses are distributed in the corpora.

4.2. Overall results

Striking differences already emerge from a comparison of the overall distribution of *MAKE* in the four corpora. The counts reported in Table 2 show that *MAKE* is most frequent in native writing (LOCNESS, with an average of 350.6 occurrences per 100,000 words), followed by learner writing (ICLE-FR, 245 occurrences per 100,000 words), leaving native speech (LOCNEC) far behind with a relative frequency of 146.8, and learner speech (LINDSEI-FR) bringing up the rear with only 126.6 occurrences per 100,000 words. However dramatic these differences may look to the naked eye, statistical measures are needed to determine if, and to what extent, these differences are statistically significant. The results were submitted to Pearson's chi-square test, and effect size is reported by means of the odds ratio (OR), as recommended by Field (2005: 693-694).[3]

	LOCNEC	LOCNESS	LINDSEI-FR	ICLE-FR
Nr of occurrences	174	528	115	394

---

[3] The chi-square test is sensitive to sample size: the larger the sample, the more likely it is that the result will be significant. An effect size is a measure that indicates the extent to which a statistically significant result is meaningful or important: it is "an objective and standardized measure of the magnitude of the observed effect" (Field 2005:32).

/100,000 words	146.8	350.6	126.6	245

Table 2: Overall results in the four corpora

In native language, there is a significant relation between the speech/writing variable and the frequency of *MAKE* (χ^2 (1)[4] = 105.962, *p* < .001): based on the odds ratio, *MAKE* is 2.4 times more likely to be produced in writing (LOCNESS) than in speech (LOCNEC). Thus, *MAKE* occurs more frequently in writing than in speech in native English.

In the learner corpora as well, *MAKE* is more frequent in writing than in speech, with approximately 126 occurrences per million words in LINDSEI and nearly twice as many in ICLE-FR. Here, too, the relation is highly significant (χ^2 (1) = 40.359, *p* < .001): according to the odds ratio, *MAKE* is 1.9 times as likely to be produced in the learner essays as in the interviews.

Comparing spoken and written production for native speakers and learners separately does not tell us much about how learners fare when compared to native speakers. Let us now therefore take the native/learner distinction as the independent variable. The two spoken corpora are very close to one another; the result is not statistically significant (χ^2 (1) = 1.521, *p* > .05). For writing, however, a highly significant association exists between the native/learner variable and the number of occurrences of *MAKE* (χ^2 (1) = 29.356, *p* < .001): based on the odds ratio, *MAKE* is 1.4 times as likely to be produced in writing by native speakers as by French-speaking learners. This means that learners significantly underuse *MAKE* in the written task, while they do not differ much from native speakers in the interviews.

One aspect of learner production which has been emphasized repeatedly is the fact that learner writing tends to resemble native speech (Granger & Rayson 1998, Crawford 2005). Learner writing seems to be situated in some kind of 'middle ground': not as typically formal as native writing, but not as informal as speech. Accordingly, it is interesting to compare the frequency of *MAKE* in learner writing (ICLE-FR) with its frequency in native speech (LOCNEC). Again, the difference in frequency is highly significant (χ^2 (1) = 32.482, *p* < .001): based on the odds ratio, *MAKE* is 1.7 times as likely to be produced in ICLE-FR as in LOCNEC. Although this result is highly significant, there are, so to speak, degrees of high significance. The contrast is much less dramatic between native speech (LOCNEC) and learner

---

[4] The number between brackets refers to the degree of freedom.

writing (ICLE-FR) than between the two native corpora. This seems to represent the fact that the essays of French-speaking learners are much closer to the native interviews in terms of the overall frequency of *MAKE* than the native essays are, lending support to the above-mentioned hypothesis of a spoken-like tendency of learner writing.

In the same way, the difference between speech and writing is more marked in native English (χ^2 (1) = 105.962, $p < .001$, OR = 2.4) than in learner language (χ^2 (1) = 40.359, $p < .001$, OR = 1.9). This indicates yet again that, in learner output, the distinction between speech and writing is less sharp than in native output: it also shows that this reduced difference is due to an underuse of *MAKE* in learner writing, pointing to a frequency in learner essays which brings them closer to native speech.

In the light of these findings, it is interesting to discuss a remark made by Borgatti (2006:75-76) in his study of the verbs *MAKE* and *DO* in the French and Dutch ICLE subcorpora. He finds it surprising that his results (namely, the fact that *MAKE* is underused by both groups of learners) should conflict with other studies of high-frequency verbs (e.g. Källkvist 1998; Ringbom 1998a) which report that learners tend to use high-frequency verbs, including *MAKE*, more often than native speakers. However, these seemingly contradictory findings come as no surprise in view of the following facts: firstly, *MAKE* behaves differently from most core verbs and is used less in speech than in writing in native English, so that its 'aberrant' behaviour does not invalidate general claims about learners' overuse of high-frequency verbs.[5] Secondly, it appears that the learners' L1 makes a difference: in his investigation of seven ICLE subcorpora, Ringbom's (1998a: 193) frequency counts show that the Swedish-speaking learners are the only group who do *not* underuse *MAKE* when compared to the native control corpus (LOCNESS) (see also Altenberg & Granger 2001). This may explain in part why many studies of high-frequency verbs do not bring out differences between *MAKE* and other high-frequency verbs: a fair number of these studies are based on data produced by Swedish-speaking students, who, unlike other learners, do not underuse *MAKE* in writing.

A study which, being concerned with a mother tongue background other than Swedish, does notice the specific status of *MAKE*, is Kaszubski's PhD thesis (2000).

---

[5] A large-scale study of high-frequency verbs in native English, carried out on the BNC and ICE-GB (the British Component of the International Corpus of English), revealed that most high-frequency verbs are significantly attracted to speech (18 out of the 25 verbs investigated), and only a few (6 out of 25, including *MAKE*) are writing-promoted (Hugon 2007).

He studies the phraseological profile of six high-frequency verbs, including *MAKE*, in the writing of four Polish-speaking learner and three native speaker groups displaying various levels of proficiency. Kaszubski notes that "[i]nterestingly, [*MAKE*] is a verb whose single-word counts do NOT testify to learner overuse" (p. 214): the highest frequency is found in the two native corpora of academic writing, whereas for the other five verbs frequency tends to rise as the level of proficiency declines. Kaszubski suggests that the "surprising" pattern of *MAKE* might be due to its "high percentage of both frozen and restricted patterns and collocations, especially with respect to writing" (p. 214). Sadly, he only looks at one side of the coin: although he remarks that speech is "the mode from which stylistic habits of Poles and other learners derive" (2000:226), he disregards differences between native speech and writing and neglects the possibility that a frequent use of *MAKE* in writing in fact indicates a proficient user – one who has mastered a number of these frozen and restricted patterns.

This first general survey of *MAKE* in the four corpora has already proved enlightening and has helped revisit previous findings from a different viewpoint. It appears that disregarding register when studying the high-frequency verb *MAKE* (and, most likely, other high-frequency items), may lead researchers to overlook the convincing hypothesis that some frequency features of learner language might be partly due to register effects and the typically speech-like nature of learner writing. In the next section, we will refine the analysis by looking at some of the many meanings exhibited by the highly polysemous *MAKE* in native and learner speech and writing.

4.3. Semantic categories

Although the overall quantitative results already provide some interesting information, they are of little use if they are not refined. *MAKE* is a polysemous verb, so speaking of underuse or overuse in general does not tell us much about how the various meanings and patterns differ across the four corpora. A qualitative analysis is needed to try and explain the general results yielded by this first quantitative overview. To quote Altenberg & Granger (2001:177), "[a] high-frequency verb such as *make* expresses a variety of meanings and enters into a whole range of structures and it is these differences in usage (...) that are of particular interest".

Cutting up the semantic space covered by a polysemous item is a very delicate task, and involves a great deal of subjectivity as to what constitutes a separate meaning, and the semantic space allocated to each of the subdivisions (see e.g. De Cock & Granger 2004). To uncover some of the similarities and differences in usage between learners and native speakers, and between the two registers, each occurrence of *MAKE* was classified manually into one of seven major semantic categories, as listed in Table 3. This classification is an adaptation of the one developed by Altenberg & Granger (2001).[6]

1. 'Core' meaning (=create, produce)	• make wings out of wax (ICLE-FR) • if I could make the laws for gun control (LOCNESS) • what light is made of (ICLE-FR)
2. Delexical uses	• when I finally make the decision (LOCNEC) • they just made a mistake (ICLE-FR)
3. Causative uses	• what makes this country great (LOCNESS) • making outcasts of the prisoners (ICLE-FR) • making her look much prettier (LOCNEC)
4. 'Money' meaning (= earn)	• making $5.00 an hour (LOCNESS) • money making is what matters (ICLE-FR)
5. Phrasal/prepositional uses	• epic that everybody made it out to be (LOCNEC) • infectious agents are made up of bacteria (LOCNESS)
6. Link verb uses (= have the right qualities for)	• would make a terrific President (LOCNESS)
7. Other conventional uses	• in order to make ends meet (LOCNESS) • two wrongs don't make a right (LOCNESS) • Madonna made it big because of her hard work (LOCNESS)
8. Unclassifiable misuses	• studies do not make more than preparing the student... (ICLE-FR) • in order to make full part of Europe (ICLE-FR) • it doesn't make her justice (LINDSEI-FR) • Union makes power (ICLE-FR)
9. Unclear from context	• the painter . first er . made her . paints her . (LINDSEI-FR)

Table 3: Major semantic uses of MAKE

4.4. Results of the classification

---

[6] It should be noted that this classification is not entirely based on meaning: phrasal verbs, for example, can hardly be considered a semantic category.

The results of the classification are given in Table 4. For each corpus, the left-hand column gives the absolute number of occurrences for the category, while the right-hand column shows the relative frequencies per 100,000 words.

	LOCNESS		LOCNEC		ICLE-FR		LINDSEI-FR	
	Raw freq	Rel freq	Raw freq	Rel freq	Raw freq	Rel freq	Raw freq	Rel freq
Core	28	18.6	12	10.1	28	17.4	**25**	27.5
Delexical	**182**	120.9	**34**	28.7	**130**	80.9	**39**	42.9
Causative	**214**	142.1	**77**	64.9	**165**	102.6	**22**	24.2
'Money'	56	37.2	5	4.2	12	7.5	0	0
Phrasal/prep.	15	10	10	8.4	17	10.6	2	2.2
Link	6	4	1	0.8	2	1.2	0	0
Other	27	17.9	**34**	28.7	23	14.3	7	7.7
Misuses	0	0	0	0	17	10.6	15	16.5
Unclear	0	0	2	1.7	0	0	5	5.5
Total	**528**	**350.6**	**174**	**146.8**	**394**	**245**	**115**	**126.6**

Table 4: Semantic classification of *MAKE*

In native writing (LOCNESS), two heavyweight categories are distinguished (in bold in Table 4): delexical and causative occurrences make up almost 75% of the total frequency (396 out of 528). For the native interviews (LOCNEC), most occurrences are spread out between three semantic categories: together, causative uses, delexical combinations and other conventional uses account for over 80% of the instances of *MAKE* (145 occurrences out of 174). Also worthy of interest is the observation that, both in native speech and in native writing, the most common meaning of *MAKE* is its causative use. This challenges Sinclair & Renouf's assertion that "[t]he primary function of *make* (...) is to carry nouns like *decision/s, discoveries, arrangements*" (1988:153). There is, they say, "textual evidence" for this, but the counts reported above show that, at the very least, this is not true of all contexts.

In French learner writing (ICLE-FR), too, the causative category has the upper hand with over 40% (165) of all the instances of *MAKE* in this corpus. The 'delexical' category follows and accounts for approximately a third of the 394 ICLE-FR occurrences (130). As for learner speech, the first three categories (core, delexical, and causative uses) make up over three quarters of the instances in LINDSEI-FR: 86 out of 115. Here, unlike in the other three corpora, it is the delexical category which is the dominant one with 39 occurrences. The core (25) and causative (22) uses run neck and neck, leaving the other, much smaller categories far behind.

Clearly, sorting the various instances of *MAKE* according to this semantic classification, simplified as it may be, already brings out a considerably more refined and nuanced picture than overall frequencies alone. Even so, the investigation can be pursued further, for instance by examining more closely the two categories that have established themselves as the most important, namely delexical and causative uses of *MAKE*. Before embarking on this in-depth analysis, however, a few paragraphs will be given over to a discussion of the various misuses of *MAKE* contained in the two learner corpora. These, in addition to over- and underuse, will provide information as to the areas learners still struggle with at an advanced level.

4.5. Erroneous uses of *MAKE*

In order to identify the erroneous instances, the following procedure was adopted. Whenever the acceptability of an occurrence of *MAKE* seemed dubious, I resorted to *the Macmillan English Dictionary* and *the Longman Dictionary of Contemporary English* to check whether one of them included it. If the use of *MAKE* could not be found listed in any of the dictionaries, I ran a search in the British National Corpus via BNC*web*, CQP Edition (Hoffmann & Evert 2006) to retrieve any potential occurrence of *MAKE* being used in the same way. If this second search proved fruitless as well, I asked a native speaker to judge the acceptability of this use of *MAKE*.

Several types of mistake could be identified in the data. The choice of the verb often poses problems: in a number of delexical occurrences, *MAKE* is used instead of another verb, e.g. **make research* instead of *do research*, **make a poll* instead of *carry out a poll* (example 5) or **make a balance* instead of *strike a balance*.

(5) If we **made** a poll among the Belgian population to know how many people... (ICLE-FR)

In the spoken data, particularly, learners tend to use *MAKE* in a rather undiscriminating way, especially when the corresponding French structure is a delexical structure involving the prototypical French counterpart of *MAKE*, i.e. *FAIRE*, e.g. **make a kind of work* (example 6)*, *make a licence dissertation*. Quite often, the noun collocate is even borrowed directly from French: e.g. **make a*

passerelle, **make a rassemblement,* or **make a stage*, where *stage* clearly stands for *teaching practice*.

(6) especially one course we have to **make** erm . a kind of work . for the end of the year (LINDSEI-FR)

Errors on the verb also appear in other semantic categories than delexical uses. In many cases, learners use *MAKE* instead of *DO* (see Borgatti 2006), as in example 7:

(7) that's the situation we have to cope with it and . well we **make** our best (LINDSEI-FR)

Sometimes it is the noun which is problematic: **make benefits* for *make profits,* or **make critics* instead of *make criticism*, or, even better, *criticize*. The latter case also illustrates another type of mistake: cases where a delexical structure is used instead of a single verb. Other, less frequent errors involve a wrong or missing article, wrong word order as in **make turn boys into men*, or wrong particle e.g. **make out a schedule*.

In a surprisingly high number of sentences, *MAKE* is used by learners with an intended meaning which is in fact not normally covered by native *MAKE*. Confusion between 'creative' *MAKE* and 'performative' *DO* has been mentioned; more striking still is the recurrent use of *MAKE* with the intended meaning of *CONSTITUTE* (example 8):

(8) this identity must be preserved because it **makes** the richness of a country (ICLE-FR)

These uses have corresponding French structures involving *FAIRE* Fr. *faire la richesse d'un pays* = literally **make the richness of a country)*, which suggests that the diverging polysemy of the two verbs might lead learners astray.

Yet another type of error involves the use of an existing lexical phrase, but used ill-advisedly and with an intended meaning which does not correspond to the actual meaning of the structure: *make a/the difference* is a case in point: French-speaking learners show a strong tendency to use it, not in the sense of 'have an effect on

someone or something, often a positive effect' (*MEDAL* 2007:408) but with the intended meaning of *make a/the distinction*. This may be due, at least partly, to a transfer from the corresponding L1 *faire la différence*[7]. This wrong use even takes precedence over the correct one: 9 of the 10 occurrences of *make a/the difference* in ICLE-FR are used to convey the meaning of *make a distinction*, as in sentence (9).

(9) Children do not **make** the difference between fiction and reality any more (ICLE-FR)

It appears that this last type of misuse has been largely overlooked in previous studies, which concentrate only on patterns containing formal errors. To my knowledge, only Nesselhauf (2004:116) notes it, when she rightly points out that "quite frequently combinations that learners use actually exist in English but are used incorrectly, i.e., they do not convey the meaning that was apparently intended". She goes on to remark that "[i]f the context of the combination is not taken into account, such mistakes are easily overlooked" (2004:117). This may help to explain the sharp contrast between the error rate of 7.2% reported by Borgatti (2006:95) on the same data[8], and my own, much higher rate of 14.7%.

Table 5 shows the number of errors for each semantic category. Unclear sentences were counted as errors in the spoken learner corpus.

	ICLE-FR		LINDSEI-FR	
	erroneous	correct	erroneous	correct
Core	3	25	2	23
Delexical	**28**	**102**	**17**	**22**
Causative	3	162	1	22
'Money'	1	11	0	0
Phrasal/prep.	4	13	1	1
Link	2	21	0	7
Other	0	2	0	0
Misuses	17	0	15	0
Unclear	0	0	5	0

---

[7] Here, L1 transfer may be reinforcing an error which seems to be widespread among learners; so much so, in fact, that the *MEDAL* devotes a 'Get it right' box to it (p. 408): "Don't use the expression **make a difference** to mean 'recognize that two things are different'. Use **make/draw a distinction**".
[8] Borgatti used the whole ICLE-FR subcorpus, including literary essays, whereas I restricted my analysis to the argumentative essays. His corpus contained 202,957 words, mine contains 160,784. This difference in the data, however, cannot account for the marked difference between the two error percentages.

Total	**58**	**336**	**41**	**74**

Table 5: Erroneous instances in ICLE-FR and LINDSEI-FR

The most striking fact brought out by these results is that delexical constructions are especially treacherous for learners: they account for over 40% of all the erroneous instances contained in the corpora (28 out of 58 and 17 out of 41 respectively). This is hardly surprising: *MAKE* is considerably depleted of meaning in such combinations, and the co-selection of the verb and its object are often a purely arbitrary and language-specific matter. In ICLE-FR, over one out of five delexical combination is erroneous (28 out of 130, or 21.5%). The proportion rises to 43.6% for speech, with 17 erroneous occurrences out of 39.

Comparatively, learner speech in general is more prone to errors than essay writing: the rate of erroneous instances is more than twice as high for LINDSEI-FR (41 out of 115 = 35.8%) as for ICLE-FR (58 out of 394 = 14.6%). As will be discussed below, the pressure of online processing is probably the main reason behind this high error rate in speech.

Now that the semantic spadework has been done on the four corpora by contrasting the spoken and written output of native and French-speaking students and identifying errors, we will concentrate in more detail on the two major semantic categories of *MAKE*: delexical and causative uses.

4.6. Delexical uses

Delexical verbs – or, more accurately, delexical *uses* of lexical verbs – have attracted a great deal of attention from linguists. There is no consensus in the literature on exactly what the term 'delexical' includes. For the present analysis, Altenberg's (2001:195) guidelines for identifying delexical uses of the verb were followed: delexical combinations "can be broadly defined as verb-object constructions in which a semantically general – or 'delexicalised' - verb is followed by a noun phrase headed by an eventive noun which carries the main part of the meaning".

There is evidence that such combinations are difficult for learners. Altenberg & Granger (2001), for example, find that both French-speaking and Swedish-speaking learners of English as a Foreign Language (EFL) underuse delexical structures. In a recent study of *MAKE*-collocations in the writing of advanced French-speaking

learners of English, Gilquin (2007) uses a multi-method approach to investigate both learners' performance and their competence, and concludes that learners are deficient at both levels when it comes to delexical constructions.

The number of delexical occurrences of *MAKE* in the four corpora is given in Table 6.

	LOCNESS	LOCNEC	ICLE-FR	LINDSEI-FR
Nr of occurrences	182	34	130	39
/100,000 words	120.9	28.7	80.9	42.9

Table 6: Delexical uses of *MAKE* in the four corpora

As mentioned before, both native and learner writing have considerably more occurrences of delexical *MAKE* than their spoken counterparts. The previous claim about delexical structures being informal therefore needs to be refined, as *MAKE* is one of the main verbs occurring in this type of structure. The contrast between the two native corpora, with a ratio of approximately 4 to 1 (120.9 vs. 28.7 occurrences per 100,000 words; χ^2 (1) = 70.287, OR = 4.22) is much more striking than it is in the learner corpora (about 2 to 1, 80.9 vs. 42.9, χ^2 (1) = 12.441, OR = 1.88).

It also appears from Table 6 that French-speaking learners do not behave consistently in their use of delexical constructions with *MAKE*: they significantly underuse them in writing, when compared to LOCNESS[9] (χ^2 (1) = 12.432, OR = 1.49), but overuse them in speech (χ^2 (1) = 2.997, OR = 1.49). Underlying this intriguing pattern are in fact two distinct causes, as revealed by an examination of the concordance lines. In writing, learners underuse typical EAP (English for Academic Purposes) structures such as *MAKE a claim/argument/statement/assumption* that are very common in native argumentative writing. This seems to result from a gap in the learners' knowledge: either they do not know these structures, or they are not aware that they play an important role in academic English.

In speech, on the other hand, it is apparently the on-line nature of oral communication that drives learners to overuse delexical *MAKE*. For one thing, as indicated earlier, the error rate for delexical combinations is very high in the learner interviews: 17 out of 39 occurrences, that is, over 40% - twice as much as for learner writing. By producing many erroneous instances in speech, learners certainly boost the frequency of this category. For another, this is a clear illustration of learners being

---

[9] This result is in accordance with the findings reported by Altenberg & Granger (2001).

confronted to what Ellis (1994:399) calls "a lack of balance between means and ends".

Clearly, this is a kind of communication strategy that learners use when they do not have the means to express what they wish: they choose a noun to express their idea and then try to find a suitable verb to attach to it. In this context, a multi-purpose verb such as *MAKE* looks like a safe bet to them; as a result, they often fall back on it and use it somewhat indiscriminately, creating infelicitous or erroneous instances such as the following:

(10) about er .. an .. experience which I .. **made** when I was in first candi (LINDSEI-FR)

This interpretation is all the more plausible if we take into account the timed/untimed variable for the ICLE-FR data: out of the 267 argumentative essays making up ICLE-FR, 253 are untimed and only 3 are timed (for the remaining 14, the answer is 'unknown'). The conditions of production are therefore diametrically opposed, from virtually no timing for the written data to complete timing for spontaneous speech.

Delexical uses of *MAKE* typically occur with collocates denoting an action being performed; among these, many express 'speech actions', i.e. something being communicated verbally, as illustrated in (11):

(11) Jefferson **made** this statement back when the federalist and democratic parties... (LOCNESS)

Altenberg & Granger (2001: 178) report an interesting tendency regarding this type of collocate: they are significantly underused in writing by both French-speaking and Swedish-speaking learners. To check whether this is mirrored in speech, all the collocates of delexical *MAKE* in the four corpora were listed manually (in lemmatized form). Table 7 is an 'abridged' version comprising only those collocates that occur at least twice. The 'speech' collocates are indicated in bold font.

LOCNESS (67 types)		LOCNEC (24 types)		ICLE-FR (45 types)		LINDSEI-FR (27 types)	
decision	31	note	5	effort	13	mistake	5
mistake	15	difference	4	difference	11	trip	4

Collocate	n	Collocate	n	Collocate	n	Collocate	n
choice	10	decision	2	progress	11	experience	3
argument	**9**	effort	2	choice	7	change	2
claim	**9**	judgement	2	mistake	7	progress	2
point	**8**			use	7	work	2
statement	**6**			discovery	6		
case	**5**			distinction	6		
difference	5			step	6		
error	5			decision	4		
effort	4			sacrifice	4		
assumption	**3**			attempt	3		
attempt	3			**point**	**3**		
impact	3			poll	3		
judg(e)ment	3			**agreement**	**2**		
advance	2			comparison	2		
appointment	2			concession	2		
calculations	2			contact	2		
call	**2**			separation	2		
contribution	2			**statement**	**2**		
demand	**2**						
discovery	2						
improvements	2						
love	2						
observation	**2**						
reference	**2**						

Table 7: Main collocates of delexical *MAKE* in the four corpora

The number of 'speech' collocates for each corpus is given in Table 8. The counts are reported in tokens, not in types. Table 8 shows that almost a third of the LOCNESS collocates belong to the 'speech nouns' category, while the other three corpora consistently display a much lower proportion of only 9-10 percent. Once again, this shows that the essays in ICLE-FR lack characteristic features of their native counterparts in LOCNESS.

	LOCNESS		LOCNEC		ICLE-FR		LINDSEI-FR	
Speech coll.	**57**	**31.5%**	3	9.4%	12	9.2%	4	10.3%
Other coll.	124	68.5%	29	90.6%	118	90.8%	35	89.7%
Total	**181**	**100%**	**32**	**100%**	**130**	**100%**	**39**	**100%**

Table 8: Speech collocates of delexical *MAKE*

Table 7 above also reveals that the nouns topping the collocate list are very different across corpora. Native writers favour above all the noun *decision*, which occurs 31

times in LOCNESS and single-handedly accounts for 17% of all delexical structures involving *MAKE* in this corpus; the list includes 26 collocates occurring at least twice. The list is much less varied in native speech, where only five types occur more than once as collocates: *note* (5), *difference* (4), *decision* (2), *effort* (2), and *judgement* (2). For learner writing, we find 20 collocates with at least two occurrences. *Decision,* the most prominent collocate of *MAKE* in LOCNESS, is conspicuously scarce in ICLE-FR with only four occurrences. This appears to be due, not to an underuse of this noun (a search on *decision(s)* yielded 26 occurrences) but rather to the fact that French-speaking learners use it more often in combination with *TAKE* (7 instances). This is most probably a case of L1 transfer, as the corresponding structure in French involves the typical translational equivalent of *TAKE*, i.e. *PRENDRE*: *prendre une decision* (*make/take a decision*). Lastly, the learner spoken corpus has 6 nouns occurring more than once as collocates of *MAKE*: *mistake* (5), *trip* (4), *experience* (3), *change* (2), *progress* (2), and *work* (2).

4.7. Causative uses

As stated before, the causative category ranks first, in terms of frequency, in native writing (LOCNESS, 40.5% of all the occurrences of *MAKE* in this corpus), native speech (LOCNEC, 44%) and learner writing (ICLE-FR, 41%), but causative uses are outnumbered by delexical uses in learner speech (LINDSEI-FR) and only account for 21% of the instances of *MAKE* in this corpus. Table 9 brings together the overall frequencies of causative *MAKE* in the four corpora. Again, unsurprisingly, this category is substantially more frequent in writing than in speech, both in native (142.1 vs 64.9 occurrences per 100,000 words; χ^2 (1) = 36.565, $p < 0.001$, OR = 2.19) and in learner production (102.6 vs. 24.2; χ^2 (1) = 48.057, $p < 0.001$, OR = 4.24). Both groups of learners considerably underuse it: for writing, χ^2 (1) = 9.973, $p < 0.01$, OR = 1.38; for speech, χ^2 (1) = 18.06, $p < 0.001$, OR = 2.68.

	LOCNESS	LOCNEC	ICLE-FR	LINDSEI-FR
Nr of occurrences	214	77	165	22
/100,000 words	142.1	64.9	102.6	24.2

Table 9: Causative uses of *MAKE* in the four corpora

Within causative uses of *MAKE*, three syntactic constructions can be distinguished: *MAKE + adjective* (*ADJ*, example 12), *MAKE + noun* (*N,* 13), and *MAKE + verb* (*V*, 14).

(12) modern man has continuously found new ways to make his life easier (LOCNESS)

(13) This would make him a less assertive person (LOCNESS)

(14) The wrong clothes can often make a person feel uncomfortable (LOCNESS)

The distribution of these structures in the four corpora is given in Table 10.

	LOCNESS		LOCNEC		ICLE-FR		LINDSEI-FR	
Verb	64	30%	42	54.5%	54	32.7%	11	50%
Adjective	122	57%	30	39%	95	57.6%	9	40.9%
Noun	28	13%	5	6.5%	16	9.7%	2	9.1%
Total	**214**	**100%**	**77**	**100%**	**165**	**100%**	**22**	**100%**

Table 10: Number of occurrences for the three causative constructions across the four corpora

The table brings out similarities, but also clear differences between the corpora. Interestingly, the rankings for the two written corpora (LOCNESS and ICLE-FR) on the one hand, and the spoken corpora (LOCNEC and LINDSEI-FR) on the other are similar: causative *MAKE* with a verb complement is the most frequent structure in the interviews, but in the essays it is the adjective construction which is most common. When one considers, not the relative frequencies of the constructions, but their respective proportions within each corpus, an interesting pattern emerges: learners seem to have a very good grasp of frequency differences within causatives. The proportions in the two written registers are remarkably close: 30% and 32.7% for the *MAKE + V* structure, 57% and 57.6% for *MAKE + ADJ*, and 13% and 9.7% for *MAKE + N*. The same can be said of the spoken data: 54.5% and 50% for *MAKE + V*, 39% and 40.9% for *MAKE + ADJ* and 6.5% and 9.1% for *MAKE + N*. This suggests that learners have a very good grasp of frequency differences within causatives, even though there is still plenty of room for improvement as advanced learners still demonstrate substantial lexical and syntactic problems in their use of causative constructions (see e.g. Gilquin forthcoming a).

5. Conclusion

MAKE has proved to be a versatile and multifaceted high-frequency verb, with an extremely diverse semantic and lexical patterning. French-speaking learners, it has been seen, use *MAKE* with an extremely varied degree of accuracy according to the semantic category. The impressive and nearly perfect match between learners and native speakers regarding the proportions of each of the three syntactic structures involving causative *MAKE*, for example, stands in sharp contrast with the largely defective use that learners make of delexical combinations, or with the virtual absence of other semantic categories in the learner data. Thus, a compelling picture gradually emerges of the scale of proficiency which seems to characterise advanced interlanguage, going from very little productive competence to near-perfect knowledge, including various levels of partial knowledge. Advanced learners, Cobb (2003:419) argues, are not "defective native speakers cleaning up a smattering of random errors"; rather, they should be seen as "working through identifiable acquisition sequences" such as lexical expansion and genre diversification.

According to Granger (2004:135), "[a]dvanced interlanguage is the result of a very complex interplay of factors: developmental, teaching-induced and transfer-related". This study brings convincing evidence of this. For instance, the tendency of learners' academic writing to resemble native speech is confirmed for French-speaking learners, and evidence from other mother tongue backgrounds corroborates this tendency (see e.g. Petch-Tyson 1998). Cobb (2003) replicates three studies carried out on European ground (Ringbom 1998b, De Cock et al. 1998, Petch-Tyson 1998) in Québec; the similarity of his results with those of the European studies is remarkable. This leads him to posit that the acquisition sequences through which advanced learners are working are "systematic and more or less universal" (Cobb 2003: 419). In other words, learners seem to have an "essentially developmental" (Granger & Rayson 1998:130) tendency to produce the same type of English irrespective of genre: "(...) learners, as a rule, fail to distinguish between genres, using a single type of English, undifferentiated between speech and writing. Put more simply, learners tend to write as they speak and speak as they write (...)" (Gilquin forthcoming b). Yet there are indications that teaching may also play a role: the current ELT methodology used in Western countries, for instance, puts much emphasis on spoken interaction, and this may reinforce learners' tendency to use

speech-like forms (Cobb 2003: 416; see also Granger & Rayson 1998). On the other hand, in Japan for instance, learners are encouraged to use written forms in all situations, as noted by Guest (1998): he finds that Japanese ELT students "often converse as if they were walking textbooks".

Some pedagogical implications arise from this study. First, we can only agree with Lennon (1996: 23) when he suggests that "teaching at the advanced level should aim (...) also to flesh out the incomplete or 'skeleton' entries which even advanced learners may have for high-frequency verbs". Cobb (2003:406) also rightly makes a case against restricting vocabulary instruction at the advanced level to "piling more and more low-frequency vocabulary into learners' heads for passive use in reading comprehension". Indeed, at this stage, teaching should aim to improve learners' productive knowledge of imperfectly acquired items, thereby fine-tuning their lexical competence. Cobb notably suggests resorting to focused awareness-raising, for instance of the formal features typifying different genres (2003:418). Gilquin likewise advises that the first step should be to raise "learners' awareness of the phraseological nature of high-frequency verbs" (2007:288). There is definitely room for pedagogical improvement in other fields too. Lexicography is a case in point: in several learners' dictionaries priding themselves on attributing frequency markers to the most common entries, either register differences or polysemy, when not both at the same time, are neatly swept under the carpet. Highly frequent words would greatly benefit from a more fine-grained approach: learners need all the help they can get when dealing with such complex items. A more adequate treatment in dictionaries, distinguishing frequency of use in (informal) speech and (formal) writing, and in the main semantic subdivisions of words with multiple meanings, for example, would be a welcome endeavour, if admittedly extremely labour-intensive.

Through the study of the verb *MAKE*, we have not only improved our knowledge of how this verb is used in native and learner speech and writing; we have also shed some light on the nature of advanced interlanguage (Cobb 2003:419) and sketched out possible pedagogical implications. If similar studies are carried out on other high-frequency verbs, we will gradually identify the acquisition sequences which advanced learners are working through.

References

Altenberg, B., 'Contrasting delexical English *make* and Swedish *göra*', in: K. Aijmer (ed.), *A Wealth of English. Studies in Honour of Göran Kjellmer*. Göteborg: Acta Universitatis Gothoburgensis, 2001, p.195-219.

Altenberg, B. & S. Granger (2001) 'The grammatical and lexical patterning of MAKE in native and non-native student writing', in: *Applied Linguistics* 22 (2001), nr. 2, p. 173-194.

Bartning, I. 'L'apprenant dit avancé et son acquisition d'une langue étrangère. Tour d'horizon et esquisse d'une caractérisation de la variété avancée', in: *Acquisition et Interaction en Langue Etrangère* 9 (1997), p. 9-50.

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

Borgatti, E., *The Use of the Verbs 'Make' and 'Do' by French- and Dutch-Speaking EFL Learners*. Unpublished MA thesis. Université catholique de Louvain, Louvain-la-Neuve 2006.

Cobb, T., 'Analyzing late interlanguage with learner corpora: Québec replications of three European Studies', in: *The Canadian Modern Language Review* 59 (2003), nr. 3, p. 393-423.

Corder, S. P., 'The significance of learner's errors', in: IRAL 5 (1967), nr. 4, p. 161-170.

Crawford, W., 'Is L2 writing like native-English conversation?' Paper presented at *ICAME 26 (International Computer Archive of Modern and Medieval English) - AAACL 6 (American Association of Applied Corpus Linguistics)*, University of Michigan, 12-15 May 2005.

de Bot, K., W. Lowie & M. Verspoor, *Second Language Acquisition. An Advanced Resource Book*. Routledge, London & New York 2005.

De Cock, S. & S. Granger, 'High Frequency Words: the Bête Noire of Lexicographers and Learners Alike. A close look at the verb 'make' in five monolingual learners dictionaries of English', in: G. Williams G. and S. Vesssier (eds.), *Proceedings of the Eleventh EURALEX International Congress*. Lorient, Université de Bretagne-Sud 2004, p. 233-243.

De Cock S., S. Granger, G. Leech & T. McEnery, 'An automated approach to the phrasicon of EFL learners', in: S. Granger (ed.), *Learner English on Computer*. Addison Wesley Longman, London & New York 1998; p. 67-79.

de Haan, P. 'How 'native-like' are advanced learners of English?', in: A. Renouf (ed.), *Explorations in Corpus Linguistics*. Rodopi, Amsterdam 1998, p. 55-66.

Doughty, C. J. & M. H. Long (eds.), *The Handbook of Second Language Acquisition*. Blackwell, Cambridge (MA) 2003.

Dulay, H. & M. Burt, 'Should we teach children syntax?', in: *Language Learning* 23 (1973), nr. 2, p. 245-258.

Dulay, H. & M. Burt, 'Natural sequences in child second language acquisition', in: *Language Learning* 24 (1974), p. 37-53.

Ellis, R., *The Study of Second Language Acquisition*. Oxford University Press, Oxford 1994.

Field, A. P., *Discovering statistics using SPSS*. Sage publications, London 2005.

Gass, S., 'Language transfer and universal grammatical relations', in: *Language Learning* 29 (1979), nr. 2, p. 327-344.

Gilquin, G. 'To err is not all. What corpus and elicitation can reveal about the use of collocations by learners', in: *Zeitschrift für Anglistik und Amerikanistik* 55 (2007), nr. 3, p. 273-291.

Gilquin, G., 'The non-finite verb slot in English causative constructions. Comparing native and learner collostructions', in: J. Leino & R. von Waldenfels (eds.), *Analytical Causatives*. Lincom Europa, München, Forthcoming a.

Gilquin, G. 'Hesitation markers among EFL learners: pragmatic deficiency or difference?', in: J. Romero-Trillo (ed.), *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Mouton de Gruyter, Berlin/ Heidelberg/ New York, Forthcoming b.

Granger S., 'Romance Words in English: from History to Pedagogy', in: J. Svartvik (ed.), *Words. Proceedings of an International Symposium.* Almqvist and Wiksell International, Stockholm 1996, p. 105-121.

Granger, S., 'The computer learner corpus: a versatile new source of data for SLA research', in: S. Granger (ed.), *Learner English on Computer*. Addison Wesley Longman, London/New York 1998 p. 3-18.

Granger, S., 'A Bird's-eye view of learner corpus research', in: S. Granger, J. Hung & S. Petch-Tyson (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Benjamins, Amsterdam & Philadelphia 2002, p. 3-33.

Granger, S., 'Computer learner corpus research: current status and future prospects', in: Connor U. and Upton T. (eds) *Applied Corpus Linguistics: A Multidimensional Perspective*. Rodopi, Amsterdam & Atlanta 2004, p. 123-145.

Granger, S., 'The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation'. To appear in: K. Aijmer (ed.), *Corpora and Language Teaching*. Benjamins, in press.

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

Granger, S. & P. Rayson, 'Automatic lexical profiling of learner texts', in: S. Granger (ed.), *Learner English on Computer*. London & New York: Addison Wesley Longman, London/New York 1998, p. 119-131.

Guest, D., 'Spoken Grammar: Easing the Transition'. 1998. Available from http://www.jalt-publications.org/tlt/files/98/jun/guest.html

Hasselgren, A., 'Lexical Teddy Bears and Advanced Learners: A Study into the Ways Norwegian Students Cope with English Vocabulary', in: *International Journal of Applied Linguistics* 4 (1994), nr.2, p. 237 258.

Hoey, M., 'Foreword', in: *Macmillan English Dictionary for Advanced Learners*. First edition. Macmillan, London 2002. Available from http://www.macmillandictionary.com/advanced_learners_dictionary_fore.htm

Hoffmann, S. & S. Evert, 'BNCweb (CQP-edition): the marriage of two corpus tools', in: S. Braun, K. Kohn & J. Mukherjee (eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Peter Lang, Frankfurt am Main 2006, p. 177-195.

Howarth, P., *Phraseology in English Academic Writing. Lexicographica*: Series maior 75. Max Niemeyer Verlag, Tübingen 1996.

Huang, J. & E. Hatch, 'A Chinese child's acquisition of English', in: E. Hatch (ed.), *Second Language Acquisition: A Book of Readings*. Newbury House, Rowley (MA) 1978, p. 118-131.

Hugon, C., *Disentangling Register and Transfer Effects in Second Language Acquisition*. Unpublished MA thesis. Université catholique de Louvain, Louvain-la-Neuve 2007.

James, C., *Errors in Language Learning and Use. Exploring Error Analysis*. Longman, London 1998.

Källkvist, M., *Form-class and Task-type Effects in Learner English. A Study of Advanced Swedish Learners*. Lund Studies in English 95. Lund University Press, Lund 1999.

Källkvist, M., 'Lexical infelicity in English: the case of nouns and verbs', in: K. Haastrup & Å. Viberg (eds.), *Perspectives on Lexical Acquisition in a Second Language*. Lund University Press, Lund 1998, p. 149-174.

Kaszubski, P., *Selected Aspects of Lexicon, Phraseology and Style in the Writing of Polish Advanced Learners of English: A Contrastive, Corpus-based Approach*. Unpublished PhD thesis. Adam Mickiewicz University, Poznan 2000.

Laufer, B., 'What's in a word that makes it hard or easy?', in: M. McCarthy & N. Schmitt (eds.), *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge University Press, Cambridge 1997, p. 140-155.

Leech, G., 'Preface', in: S. Granger (ed.), *Learner English on Computer*. Addison Wesley Longman, London/ New York 1998, p. xiv-xx.

Lennon, P., 'Error: Some Problems of Definition, Identification, and Distinction', in: *Applied Linguistics* 12 (1991°, nr.2, p. 180-196.

Lennon, P., 'Getting "easy" verbs wrong at the advanced level', in: *IRAL* 34 (1996), nr.1, p. 23-36.

Levenston, E. A. & S. Blum, 'Aspects of Lexical Simplification in the Speech and Writing of Advanced Adult Learners', in: S. P. Corder & E. Roulet (eds.), *Actes du 5ème Colloque de Linguistique Appliquée de Neuchâtel. 20-22 mai 1976: The Notions of Simplification, Interlanguages and Pidgins and their Relation to Second Language Pedagogy*. Droz, Genève 1976, p. 51-71.

Longman Dictionary of Contemporary English. 3rd edition. Harlow: Longman, Harlow 2003.

Macmillan English Dictionary for Advanced Learners. 2nd edition. Macmillan, Oxford 2007.

McLaughlin, B., *Theories of second-language learning.* London: Edward Arnold, London 1987.

Milon, J., 'The development of negation in English by a L2 learner'. *TESOL Quarterly* 8 (1974), p. 137-143.

Nation, I. S. P., *Teaching and Learning Vocabulary.* Newbury House, New York 1990.

Nesselhauf, N., 'How learner corpus analysis can contribute to language teaching: A study of support verb constructions', in: G. Aston, S. Bernardini and D. Stewart (eds.), *Corpora and Language Learners.* John Benjamins, Amsterdam/ Philadelphia 2004, p. 109-124.

Nesselhauf, N., *Collocations in a Learner Corpus.* Benjamins, Amsterdam 2005.

Petch-Tyson, S., 'Reader/Writer Visibility in EFL Persuasive Writing', in: S. Granger (ed.), *Learner English on Computer.* London & New York: Addison Wesley Longman, London/ New York 1998, p. 107-118.

Ringbom, H., 'High-frequency verbs in the ICLE corpus', in: A. Renouf (ed.), *Explorations in Corpus Linguistics.* Rodopi: Amsterdam/ Atlanta 1998a, p. 191-200.

Ringbom H., 'Vocabulary frequencies in advanced learner English: A cross-linguistic approach', in: S. Granger (ed.), *Learner English on Computer.* Addison Wesley Longman, London & New York 1998b, p. 41-52.

Scott, M., *Wordsmith Tools 4.0.* Oxford University Press, Oxford 1999.

Selinker, L., 'Interlanguage', in: *IRAL* 10 (1972), nr.3, p. 209-231.

Sinclair, J., *Corpus, Concordance, Collocation.* Oxford University Press, Oxford 1991.

Sinclair, J. & A. Renouf, 'A lexical syllabus for language learning', in: R. Carter & M. McCarthy (eds.), *Vocabulary and Language Teaching.* Longman, London/ New York 1988, p. 140-158.

Van Roey, J., *French-English Contrastive Lexicology: An Introduction.* Peeters, Louvain-la-Neuve 1990.

Viberg, Å., 'Cross-linguistic lexicology: The case of English go and Swedish gå', in: K. Aijmer, B. Altenberg & M. Johansson (eds.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4-5 March 1994.* Lund University Press, Lund 1996, p. 151-182.