Lexical infelicity in causative constructions.
Comparing native and learner collostructions

Gaëtanelle Gilquin

F.N.R.S.

Université catholique de Louvain

Abstract

This paper seeks to describe foreign learners' use of English periphrastic causative constructions, by comparing data from the International Corpus of Learner English *with similar data from the* British National Corpus. *While the frequency of the constructions and the syntactic errors found in the learner corpus are briefly discussed, the focus of the study is on collostructions, the combination of particular constructions with certain verbs in the non-finite verb slot. After showing that the non-finite verb slot displays strong lexical preferences in native English, the technique of distinctive collexeme analysis is used to compare native speakers' and learners' lexical choices and quantify the degree of overlap between the two groups. The analysis reveals that learners' causative constructions are often unidiomatic in that they contain verbs which are unlikely to occur in native English with that particular construction. Three possible explanations are provided to account for these lexical infelicities, namely lack of register awareness, transfer from the mother tongue and inadequacy of teaching materials. The article concludes by emphasising the necessity to explicitly teach phraseological aspects of language.*

1. Introduction

Most learners of a foreign language will be familiar with the experience of being told that a particular sentence is perfectly grammatical in the language but that a native speaker would never use it (cf. Allerton 1984: 39). Such problems belong to the field of phraseology, that is, "the study of word combinations" (Howarth 1998: 24). Thanks to the advent of corpus linguistics and the use of large collections of authentic texts searchable at the click of a mouse, the importance of prefabricated units and lexical preferences (collocations) has been

widely recognised. It has been shown that "much language use is routine" (Stubbs 1993: 19) and that, although they may only be perceived at a subliminal level (Sinclair 1991: 116), collocations are an integral part of language.

This paper examines the phraseology of English periphrastic causative constructions and the problems it poses to foreign learners. More precisely, it investigates the lexical preferences displayed by the non-finite verb slot of the ten constructions listed and exemplified in Table 1, with particular emphasis on [X *MAKE* Y V_{inf}] and [X *MAKE* Y V_{pp}]. By means of a technique known as distinctive collexeme analysis, it compares the lexical choices made by learners from various mother tongue backgrounds with those of native speakers of English, highlighting the verbs that are more distinctive for learner writing or more distinctive for native writing. The analysis shows that learners have difficulty producing combinations of a construction and a non-finite verb (so-called "collostructions") that are likely to occur in native language, which results in grammatically correct, but unidiomatic causative constructions. This problem appears to be at least as important as the problem of syntactic inaccuracy (a problem regularly acknowledged in the literature; cf. among others the error note on *make* in the *Longman Dictionary of Common Errors*, Turton & Heaton 1996) or that of over- or underuse (see section 3 on these two notions), which will also be briefly examined in the analysis.

[X *MAKE* Y V_{inf}]	This **made** the accident *appear* reasonable, something which even they could have done. <BNC:A5Y 1310>
[X *BE made* $V_{to\text{-}inf}$]	It was usually supposed that teaching should follow the order of discovery, for then the student is **made** *to begin* with familiar things. <BNC:ABM 138>
[X *MAKE* Y V_{pp}]	Shortages of foodstuffs and consumer goods were also **making** themselves *felt*. <BNC:ADD 358>
[X *GET* Y $V_{to\text{-}inf}$]	British Prime Ministers may have to work extremely hard to **get** them [the departments] *to change* course. <BNC:A6F 723>
[X *GET* Y V_{pp}]	They are often an inaccessible but important target group in agricultural extension efforts to **get** conservation methods *accepted*. <BNC:APN 1038>
[X *GET* Y V_{prp}]	Yet there is little doubt that in most countries a good deal more could be done to **get** people *talking* and *thinking* about proposed changes. <BNC:BLY 1742>
[X *HAVE* Y V_{inf}]	In my own research into books and reading I have **had** classes of 15-year-olds *write* essays on the subject of how they would feel about working in a bookshop. <BNC:B25 599>
[X *HAVE* Y V_{pp}]	Eventually we end up with samples of people who have **had** children *baptized* at two Anglican churches. <BNC:B25 246>
[X *HAVE* Y V_{prp}]	After the day's work we **had** all the colony *drilling* for an hour or two in

	the yard, which formed a spacious square. <BNC:A64 952>
[X *CAUSE* Y V*to*-inf]	As the central circle rotates, it **causes** the rolling circle *to give* point P a reciprocating action. <BNC:ADX 66>

Table 1. English periphrastic causative constructions[1]

The paper is organised as follows. In section 2, it is demonstrated that, contrary to a common assumption, English periphrastic causative verbs display strong lexical preferences in terms of the non-finite verb with which they occur. The existence of such "felicitous" choices opens the door to "infelicitous" choices, once constructions are used with unusual verbs. Section 3 presents the technique of collostructional analysis, and in particular distinctive collexeme analysis, which makes it possible to measure the degree of resemblance between native speakers' and learners' lexical choices. It also describes the learner corpus used in the study and the native corpus serving as a baseline. The following section reports the main results of the corpus analysis, briefly considering the frequency of the constructions in native and learner English and the syntactic errors made by learners, and focusing on the collostructions distinctive for the two varieties of English. In section 5, three possible explanations are provided to account for learners' lack of awareness of the lexical preferences displayed by causative constructions. Section 6, finally, offers some concluding remarks.

2. Lexical preferences in causative constructions

While it is usually assumed that periphrastic causative constructions are "always safe" (Stocker 1990: 61) and can therefore be used with any verb in the non-finite verb slot, a close look at authentic data of native English reveals that periphrastic causative verbs actually display strong preferences for certain verbs. Thus, a construction such as (1), where *make* is construed with the infinitive *feel*, is more likely to occur than (2), where the infinitive *talk* is used.

(1)   That's right, if it **makes** you *feel* good inside I think you should go for it. <BNC:FL8 212>

---

[1] The examples in Table 1 are native sentences extracted from the *British National Corpus* (see later on this corpus). In these and the following examples, the causative is in bold and the non-finite verb in italics. The code between angle brackets is the reference of the sentence in the corpus, *British National Corpus* (BNC) or *International Corpus of Learner English* (ICLE).

(2)   He always **makes** people (…) *talk* about these mad things that they probably

        wouldn't talk about that often. <BNC:KC7 516>

Tables 2 to 4, taken from Gilquin (2004), show the absolute and cumulative frequencies of the non-finite verbs in [X *MAKE* Y V_{inf}], [X *GET* Y V_{prp}] and [X *MAKE* Y V_{pp}], respectively, as evidenced in a written and spoken subset of the *British National Corpus, World Edition* (Burnard 2000). With [X *MAKE* Y V_{inf}], the most frequent verb, *feel*, accounts for over 12% of all the occurrences of the construction, and a cumulative frequency of 50% is reached after nine verbs only. If we consider [X *GET* Y V_{prp}], we see that the lexical preferences are even stronger, with the verb *going* representing almost 57% of the data. With [X *MAKE* Y V_{pp}], finally, the top verb, *known*, accounts for as many as 60% of all the occurrences of the construction, and the top three verbs, *known*, *felt* and *understood*, represent a cumulative frequency of some 83%.

	Abs. freq.	Cum. Freq.
Feel	12.29%	12.29%
Laugh	9.44%	21.73%
Look	7.45%	29.18%
Think	5.71%	34.89%
Go	3.98%	38.87%
Do	3.90%	42.77%
Wonder	2.68%	45.45%
Appear	2.51%	47.96%
Seem	2.42%	50.38%
Work	2.25%	52.63%

Table 2. Absolute and cumulative frequencies of the non-finite verbs in [X *MAKE* Y V_{inf}] (based on Gilquin 2004)

	Abs. freq.	Cum. Freq.
Going	56.72%	56.72%
Doing	3.73%	60.45%
Running	3.73%	64.18%

Table 3. Absolute and cumulative frequencies of the non-finite verbs in [X *GET* Y V_{prp}] (Gilquin 2004)

	Abs. freq.	Cum. Freq.
Known	60.00%	60.00%
Felt	16.67%	76.67%
Understood	6.67%	83.34%

Table 4. Absolute and cumulative frequencies of the non-finite verbs in [X *MAKE* Y V_{pp}] (Gilquin 2004)

Not only do periphrastic causative constructions display strong lexical preferences, but they also display lexical preferences which are largely specific to each construction (see Gilquin 2006). [X *HAVE* Y V_{pp}], for example, favours verbs from the frame of service (e.g. *build*, *shave* or *repair*), whereas [X *GET* Y V_{pp}] is often associated with verbs of organisation like *sort out*, *organise* or *set up*. Constructions with the same causative verb but distinct patterns tend to display different preferences too. Thus, unlike [X *GET* Y V_{pp}], [X *GET* Y V_{prp}]

is strongly associated with verbs of (literal or metaphorical) motion (*go*, *run*, *move*, etc). These preferences are illustrated by the following examples:

> (3) He's **had** his hair *cut* this morning, reckons on having it cut short, really short. <BNC:KCE 3494>
>
> (4) Cos I want to try and **get** things *sorted out* <pause> round there. <BNC:KBE 3793>
>
> (5) We **get** a game of cards *going* just to stop the rot and the boredom. <BNC:KCP 4712>

The fact that such lexical preferences are often ignored in the literature[2] and can only be brought to light by a careful examination of authentic language data seems to suggest that they escape speakers' conscious attention. They are therefore challenging for foreign learners, who, in their search for felicitous lexical choices, can neither rely on their intuitions, nor on appropriate reference tools. In the next section, we will see how the method of collostructional analysis can be used to assess the extent to which learners approximate to native speakers' choices in the non-finite verb slot of causative constructions.

3. Collostructional analysis and learner corpora

The method of collostructional analysis, developed by Gries and Stefanowitsch and relying on the notion of collocation and the theory of Construction Grammar,[3] seeks to investigate the interaction between words and constructions, as evidenced in corpus data, by measuring the association strength that exists between a particular construction and the lexemes occurring in a given slot of this construction, i.e. its "collexemes" (see Stefanowitsch & Gries 2003, Gries & Stefanowitsch 2004a, b). While collostructional analysis includes different techniques, it is the technique of distinctive collexeme analysis that interests us here. Originally, distinctive collexeme analysis studies one slot in two or more constructions. Thus, Gries & Stefanowitsch (2004a) employed this technique to compare the verbs occurring in the

---

[2] The preferences of [X *MAKE* Y V$_{pp}$] are a notable exception. See e.g. Van Ek and Robat (1984: 327), who note that the construction occurs with verbs "denoting the exercise and recognition of influence in the widest sense".

[3] Construction Grammar considers that constructions are the basic units of language and that they carry a meaning of their own, independently of the words they are made up of. For more information on this theoretical framework, see Goldberg (1995, 2006) and, for an acquisitional approach, Tomasello (2003).

ditransitive construction with those occurring in the *to*-dative construction, and in Gilquin (2006, in preparation) it was used to establish the collexemes that are most distinctive for the non-finite verb slot of different periphrastic causative constructions. For the present purposes, however, distinctive collexeme analysis was applied to the comparison, not of two constructions, but of one construction (each of the ten periphrastic causative constructions listed in Table 1) in two language varieties, namely native English and learner English. This way, it was possible to identify the verbs that are more distinctive for one or the other variety of English in a given causative construction, in other words, the "collostructions" (i.e. combinations of a construction and a collexeme) favoured by native speakers and neglected by learners, and vice versa.

The analysis was performed by means of *Coll.analysis 3* (Gries 2004) which, on the basis of a list of the verbs occurring in the construction(s) under investigation, computes the observed and expected frequencies and, through a binomial test, the (log-transformed) probability of a particular observed frequency given the expected frequency. The "distinctiveness value" resulting from this calculation indicates which verbs are significantly associated with a given construction, that is, are distinctive collexemes of this construction.[4] In the present case, the higher the distinctiveness value, the more distinctive the verb is for the construction in a particular language variety (native or learner). The threshold level of significance is 1.30103, which means that any distinctiveness value higher than 1.30103 is statistically significant, with *p*<0.05.

Distinctive collexeme analysis (or any type of collostructional analysis, for that matter) presupposes the use of authentic language data. While corpora of native language have been in use for quite some time now (the Brown Corpus, for example, was started in the early 1960s), learner corpora, that is corpora made up of texts produced by language learners, are more recent, with the first academic learner corpus, the *International Corpus of Learner English* (ICLE), released in 2002 (Granger et al. 2002). Yet, the numerous studies that have analysed data from this and other learner corpora over the last few years have demonstrated the usefulness of such corpora, both on their own and in comparison with native corpora. Not only do they give access to authentic learner errors in context but, when exploited in conjunction with a reference corpus of native language, they also make it possible to highlight cases of over- and underuse, that is, cases where learners use significantly more or significantly less of a particular phenomenon than native speakers (e.g. underuse of the

---

[4] For more information on the computation of distinctive collexeme analysis, see the help files accompanying *Coll.analysis 3* (Gries 2004) or, for concrete applications, Gries & Stefanowitsch (2004a) or Gilquin (2006).

passive voice or overuse of amplifying adverbs). In addition, by comparing corpus data produced by learners with different mother tongues, one can distinguish between problems that are specific to learners from a given mother tongue background (and are therefore likely to be transfer-related) and those that are common to most or all learners, whatever their mother tongues (and are therefore more likely to be developmental).

This study is based on data from the second version of ICLE (Granger et al. forthcoming), consisting of academic essays written by learners from sixteen different mother tongue backgrounds,[5] for a total of over 3.5 million words. The causative constructions were extracted by means of *WordSmith Tools 3* (Scott 1999), using a combination of automatic extraction (a form of the causative verb followed, within one to five words, by an infinitive, past participle or present participle)[6] and manual post-editing (in order to discard irrelevant hits). Because some causative constructions are very infrequent and because collostructional analysis (like any type of phraseological analysis) requires a sufficient amount of data for its results to be relevant, no distinction was made between the different learner populations and the possible influence of the mother tongue was therefore not investigated. The learner data were compared with native data extracted from part of the academic component of the *British National Corpus, World Edition* (Burnard 2000), henceforth BNCw, consisting of 5 million words equally distributed over the different genres of the academic component (humanities, medicine, natural sciences, politics/law/education, social & behavioural sciences, and technology/computing/engineering). The search strings were similar to those used in ICLE,[7] but the text retrieval software employed was *BNCweb* (Lehmann et al. 2002). Although the focus of the analysis was on the collostructions distinctive for learner English, as opposed to native English, the frequencies of the different constructions were also compared in the two corpora, and the syntactic errors made by learners were examined. The main results are presented in the next section.

4. Causative constructions in learner English

---

[5] The sixteen mother tongue backgrounds are: Bulgarian (BU), Chinese (CH), Czech (CZ), Dutch (DU), Finnish (FI), French (FR), German (GE), Italian (IT), Japanese (JP), Norwegian (NR), Polish (PO), Russian (RU), Spanish (SP), Swedish (SW), Tswana (TSW) and Turkish (TU).

[6] Note that the search strings allowed for the retrieval of non-standard patterns, such as the use of a present participle with *cause* ([X *CAUSE* Y V_{prp}]) or the positioning of the causee after the non-finite verb slot (e.g. [X *MAKE* V_{inf} Y]). In what follows, these constructions are classified together with their standardised variants, e.g. [X *CAUSE* Y V_{prp}] with [X *CAUSE* Y $V_{to\text{-}inf}$] and [X *MAKE* V_{inf} Y] with [X *MAKE* Y V_{inf}].

[7] Only the number of words between the causative verb and the non-finite verb was slightly larger (up to six words), as a pilot study showed the average distance to be longer in native English.

4.1. Frequencies

	Rel. freq. BNCw (per million w.)	Rel. freq. ICLE (per million w.)	X²	Overuse (+), underuse (-) or non-significant (ns)
[X *MAKE* Y V_{inf}]	50.77	403.93	1316.54	+
[X *BE made* $V_{to\text{-}inf}$]	14.99	14.91	0.001	ns
[X *MAKE* Y V_{pp}]	4.40	10.49	11.22	+
[X *GET* Y $V_{to\text{-}inf}$]	9.79	5.25	5.51	-
[X *GET* Y V_{pp}]	3.80	7.73	5.96	+
[X *GET* Y V_{prp}]	0.60	1.10	0.66	ns
[X *HAVE* Y V_{inf}]	5.40	2.76	3.40	ns
[X *HAVE* Y V_{pp}]	8.99	13.25	3.53	ns
[X *HAVE* Y V_{prp}]	0.60	2.76	6.51	+
[X *CAUSE* Y $V_{to\text{-}inf}$]	36.98	32.30	1.31	ns

Table 5. Relative frequency of causative constructions in BNCw and ICLE

A comparison of the frequencies of the different causative constructions in native and learner English reveals some significant differences. The most striking one concerns [X *MAKE* Y V_{inf}], which is highly significantly overused by learners ($X^2 = 1316.54$, $p<0.001$). This result confirms studies such as those by Wong (1983) or Liu & Shaw (2001), who note an overuse of the *make* causative construction among Chinese-speaking learners, or Altenberg & Granger (2001), who found such an overuse in Swedish-speaking learners' production.[8] One may hypothesise, as Wong (1983: 152) does for Chinese learners, that the overuse of [X *MAKE* Y V_{inf}] is due to learners' tendency to fall back on *make* when they need to express causation, to the detriment of other devices. This, however, would imply that the other causative constructions are all underused by learners. But as appears from Table 5, only the infinitive construction with *get* is significantly underused ($X^2 = 5.51$, $p<0.05$). By contrast, three other constructions are overused, namely [X *MAKE* Y V_{pp}] ($X^2 = 11.22$, $p<0.001$), [X *HAVE* Y V_{prp}] ($X^2 = 6.51$, $p<0.05$) and [X *GET* Y V_{pp}] ($X^2 = 5.96$, $p<0.05$), and the remaining constructions do not display any significant difference in frequency between BNCw and ICLE.

4.2. Syntactic errors

---

[8] Altenberg & Granger (2001) found a (non-significant) underuse of the construction among French-speaking learners. However, it should be underlined that they used a corpus of novice writing as their reference native corpus (*Louvain Corpus of Native English Essays*, made up of essays written by American and British university students). This is to be contrasted with the reference corpus chosen here, BNCw, which represents expert writing and in comparison to which all learner populations overuse [X *MAKE* Y V_{inf}].

The data from ICLE also present a number of syntactic errors. Among these, a distinction may be drawn between errors related to the non-finite verb and errors having to do with the causee. The former are illustrated by (6), where *make* is used with a *to*-infinitive, or (7), where *get* is followed by a bare infinitive. The latter are found in examples like (8), where the causee is misplaced, or (9), where it is missing.

(6) It is good practice to use English every day. That **makes** us not *to forget* English grammar or words. <ICLE-JP>

(7) To sum up, doing everything which could contribute to **get** our university degree *be* really worthy of its name. <ICLE-SP>

(8) Compensation orders could **make** *feel* <u>parents</u> more responsible for their sons and daughters. <ICLE-IT>

(9) It **causes** *to lose* their good emotions such as love, mercy and the other good ones. <ICLE-TU>

Some learner constructions combine these two problems (which explains why the two columns in Table 6 below do not necessarily add up), as exemplified by the following two sentences:

(10) Since the first hints of the birth of industrialization man has been inmersed in a world of new and fascinating technologies that have **made** <u>of him</u> *to be* more and more apart from the world he comes from. <ICLE-SP>

(11) It may reduce the free time between the couples, **causes** *losing* the sweet and roma[n]tic feeling. <ICLE-CH>

In (10), the wrong choice of complement (*to*-infinitive instead of bare infinitive) is combined with the wrong form of the causee (*of* + NP instead of simple NP). In (11), *cause* is used with a present participle and the causee is missing. While it is not always easy to determine whether the learner really had the causative pattern in mind when producing sentences such as these (in [11], for example, s/he may have had the transitive pattern of *cause* in mind), the fact that the two types of errors occur individually (cf. absence of the causee in [9] and use of a present participle with *cause* in [12]) suggests that the causative interpretation is tenable.

(12)   The ease payment of credit card **cause** the students *buying* too much and too quickly. <ICLE-CH>

	Errors with verb	Errors with causee	Total
CAUSE	13.68%	14.53%	**23.93%**
MAKE	14.07%	0.77%	**14.66%**
HAVE	5.88%	0.00%	**5.88%**
GET	3.92%	0.00%	**3.92%**

Table 6. Proportion of syntactic errors in learner English (ICLE)

Table 6, which gives the proportion of errors with respect to the overall number of causative constructions with the specific verb, shows that the highest number of errors is found with *cause*, followed by *make*. Proportionally, errors in *get* and *have* constructions are not so common. Except for *cause* constructions, where the two types of errors, verb-related and causee-related, represent similar percentages, errors with the non-finite verb are more frequent than errors with the causee. As pointed out by Wong (1983: 153), it is to a large extent an arbitrary matter which non-finite verb is permitted in which construction,[9] so that learners are likely to have some difficulty with the choice of an appropriate form. Particularly common is the confusion between *to*-infinitives and bare infinitives (77% of all verb-related errors),[10] but the erroneous use of a present participle also occurs (22% of all verb-related errors), especially in *cause* constructions, where all the verb-related errors are of this type. As for causee-related errors, they are predominantly due to the absence of the causee (52% of all causee-related errors) and, somewhat less frequently, to the postposition of the causee (41% of all causee-related errors), as exemplified by (9) and (8) above, respectively. Notably, the postposition of the causee is often found in the Italian and Spanish components of ICLE, which is probably due to the fact that in Romance languages, the causee normally comes after the non-finite verb (cf. Italian *Maria fa scrivere Gianni*, lit. "Mary makes write Johnny").

4.3. Collostructions

---

[9] Some authors, however, have tried to find a rationale behind these complementation patterns, see e.g. Adamczewski & Delmas (1982: 13ff) on the difference between *to*-infinitives and bare infinitives in causative constructions.

[10] The use of a *to*-infinitive with active *make* accounts for some 80% of all verb-related errors with *make*, which confirms the relevance of the error note found in the *Longman Dictionary of Common Errors* (Turton & Heaton 1996).

In addition to syntactic errors such as those described in the preceding section, the data from ICLE contain constructions which, though syntactically correct, somehow sound non-native-like. Consider (13) to (16) below. In (13), the correct idiom is *make (both) ends meet*, not *have both ends meet*. (14) illustrates a case of decomposition (see Altenberg & Granger 2001), with the learner using a causative construction where a native speaker would prefer a more synthetic alternative (here, the lexical causative verb *support*). In (15), the verb *help* would sound more natural than the verb *cause*, and in (16), *make their products known* would be a more normal option.

(13)   The appearance of "new Russians" has given a terrible imprint on common people, who lead a quiet life, earn their living at establishments and **have** both ends *me[e]t*. <ICLE-RU>

(14)   They just want to earn money and **make** the family *live*. <ICLE-FR>

(15)   First, reducing smokers to smoke in restaurants can make the environment in restaurants and bars better. This can **cause** us *[to] avoid* (…) to breath secondhand smoke and be threaten by the bad environment to affect our health. <ICLE-CH>

(16)   Secondly, many companies like to advertise their products on the television because they find it easy to **get** their products *known* to everybody as people of different ages watches television. <ICLE-CH>

While such lexical infelicities have on occasion been dealt with in the literature (cf. Wong 1983, Altenberg & Granger 2001 or Liu & Shaw 2001), it is in the form of scattered observations rather than a comprehensive account of learners' lexical problems when using causative constructions. One reason for this is certainly that lexical infelicities are thought to be, in Wong's (1983: 160) words, "not amenable to simple quantification". Using the technique of distinctive collexeme analysis, however, and applying it to the comparison of native and learner corpora, it is possible, precisely, to identify and quantify the lexical idiosyncrasies found in learners' causative constructions.[11]

Construction	Proportion of

---

[11] It should be pointed out that the distinctive collexeme analysis carried out here is based on the occurrences of the non-finite verbs, not the causative verbs. Thus, a construction like (i), which contains two non-finite verbs (*supply* and *cooperate*), will count as two in the collostructional analysis.

    (i) The great affair not to relate, but to **make** the reader *supply*, *cooperate*. <BNC A6B 1107>

	shared collexemes
[X *MAKE* Y V_{pp}]	20.00%
[X *GET* Y V_{prp}]	20.00%
[X *BE made* $V_{to\text{-}inf}$]	18.29%
[X *MAKE* Y V_{inf}]	16.34%
[X *GET* Y V_{pp}]	11.11%
[X *CAUSE* Y $V_{to\text{-}inf}$]	9.27%
[X *HAVE* Y V_{inf}]	9.09%
[X *GET* Y $V_{to\text{-}inf}$]	7.84%
[X *HAVE* Y V_{pp}]	3.80%
[X *HAVE* Y V_{prp}]	0.00%

Table 7. Proportion of collexemes in the non-finite verb slot

shared by BNCw and ICLE (%)

As a rough approximation of the overlap between the collostructions used by native speakers and those used by learners, Table 7 shows the proportion of collexemes in the non-finite verb slot of causative constructions shared by native speakers (BNCw) and learners (ICLE). It emerges from this table that the degree of overlap is relatively low. The highest proportion is with [X *MAKE* Y V_{pp}] and [X *GET* Y V_{prp}] and amounts to 20% only. Half of the constructions display a percentage of shared collexemes of under 10%, and with [X *HAVE* Y V_{prp}], no collexemes at all are shared by native speakers and learners.

4.4. *Make* and its collexemes

In what follows, we will specifically focus on causative *make*, and more particularly, on the collexemes of [X *MAKE* Y V_{inf}] and [X *MAKE* Y V_{pp}]. Table 8 lists the collexemes of [X *MAKE* Y V_{inf}] that are significantly more distinctive for native English (upper part of the table) and those that are significantly more distinctive for learner English (lower part). The observed and expected frequencies are given, together with the distinctiveness value of the collexeme.

Collexemes	Obs. freq. BNCw	Obs. freq. ICLE	Exp. freq. BNCw	Exp. freq. ICLE	Distinc-tiveness	Preferred variety
seem	25	9	4.93	29.07	14.2331	BNCw
appear	23	7	4.35	25.65	13.7779	BNCw
refer	5	1	0.87	5.13	3.4853	BNCw
work	14	26	5.80	34.20	3.0650	BNCw
vanish	5	2	1.01	5.99	2.9959	BNCw
conform	3	0	0.43	2.57	2.5207	BNCw
ask	2	0	0.29	1.71	1.6790	BNCw
fire	2	0	0.29	1.71	1.6790	BNCw
jump	2	0	0.29	1.71	1.6790	BNCw

lie	2	0	0.29	1.71	1.6790	BNCw
submit	2	0	0.29	1.71	1.6790	BNCw
sound	3	2	0.72	4.28	1.6192	BNCw
look	10	26	5.22	30.78	1.5673	BNCw
be	1	62	9.13	53.87	3.2833	ICLE
become	1	44	6.52	38.48	2.1547	ICLE
believe	1	44	6.52	38.48	2.1547	ICLE
go round	0	25	3.62	21.38	1.7127	ICLE
lose	0	24	3.48	20.52	1.6437	ICLE
feel	21	191	30.73	181.27	1.6205	ICLE
come	3	52	7.97	47.03	1.5091	ICLE
have	0	22	3.19	18.81	1.5058	ICLE
forget	0	20	2.90	17.10	1.3680	ICLE

Table 8. Distinctive collexemes of [X *MAKE* Y V_{inf}] in BNCw and ICLE

(significant values only)

What is striking in the upper part of the table is that it contains four copular verbs, viz. *seem*, *appear*, *sound* and *look*,[12] as illustrated by (17) and (18). These verbs belong to the functional category of relational verbs, as opposed to mental and material verbs (see Halliday 2004 on this threefold distinction), which seems to confirm Altenberg & Granger's (2001: 183) point that learners tend to underuse relational verbs with causative *make* (they were referring to Swedish- and French-speaking learners).

(17)   For families who can afford to pay the fees, the poor record of state schools **makes** private education *seem* attractive, even if in reality its quality is poor. <BNC:B12 1322>

(18)   There are many aids to **make** plots *look* more professional. <BNC:B16 795>

However, the situation is slightly more complex than that. It is not so much that learners underuse relational verbs, but rather, they use relational verbs which are unlikely to be used by native speakers. This transpires from the lower part of Table 8, which shows that verbs like *be*, *become*, *feel* or *have*, which are all relational verbs, are distinctive for learner English.[13] In native English, on the other hand, these verbs are relatively rare (one occurrence

---

[12] While a verb such as *look* may also be used as a non-copular verb (e.g. *look through the window*), it is almost always used as a copula in the *make* causative construction (see Gilquin in preparation on the interplay between verb senses and collexemes).

[13] Liu & Shaw (2001: 179), using a simple frequency-based approach, note a similar predilection for *feel* and *become* among Chinese-speaking learners. While they attribute this behaviour to the influence of the mother tongue, pointing out that [X *MAKE* Y *feel*] and [X *MAKE* Y *become*] have word-for-word equivalents in Chinese, it looks as if a more general tendency is at work here, since the former collostruction is found in all the components of ICLE and the latter, in ten components.

of *be* and *become*, and no occurrences of *have*). A close examination of the sentences produced by learners with one of these verbs reveals their unidiomatic nature. Consider the following examples:

(19) He passed through hardships and sufferings in order to **make** their dreams *become* true. <ICLE-RU>

(20) That will **make** it *be* more popular. <ICLE-TSW>

(21) This **made** women *become* increasingly aware of their rights. <ICLE-NR>

(22) This change also **makes** auditors *be* in a difficult situation. <ICLE-JP>

In (19), *make their dreams **come** true* would be the normal expression. In (20) and (21), the non-finite verb is redundant and a native speaker would probably use an adjectival construction instead (*make it more popular*, *made women increasingly aware*). This "preference for verbosity", highlighted by Liu & Shaw (2001: 180) for Chinese-speaking learners, partly explains learners' general overuse of [X *MAKE* Y V$_{inf}$]. It also seems to go hand in hand with an underuse of the adjectival (and nominal) causative construction among certain learner populations (see Altenberg & Granger 2001 for French-speaking learners and Liu & Shaw 2001 for Chinese-speaking learners). As for (22), native speakers may find the sentence more natural with the verb *cause*. This is because, as appears from a distinctive collexeme analysis of all ten causative constructions in native English, the verb *be* (and *become*) is more distinctive for [X *CAUSE* Y V$_{to-inf}$] than for any other construction (see Gilquin 2006).[14]

Another verb that figures prominently among the most distinctive collexemes of [X *MAKE* Y V$_{inf}$] in learner writing is *believe*, as in (23). While this verb is possible in native writing, it is not very frequent (just one occurrence in BNCw). Again, it is a verb that is more distinctive for another causative construction, namely [X *HAVE* Y V$_{inf}$], as exemplified by (24).[15] When used in [X *MAKE* Y V$_{inf}$], it is normally in the (causeeless) expression *make believe*, meaning "to pretend", cf. (25).

---

[14] [X *CAUSE* Y V$_{to-inf}$] is the only construction that has a relation of attraction, rather than repulsion, with *be* and *become*. This attraction is statistically significant, and *be* and *become* are the first and third most distinctive collexemes of the construction, respectively.
[15] *Believe* is the third most distinctive collexeme of [X *HAVE* Y V$_{inf}$], with a significant distinctiveness value of 3.53 (Gilquin 2006).

(23)   The socialists **made** us *believe* in the close bound between higher education and success in life. <ICLE-BU>

(24)   The bulletin is not a wide ranging, objective, scientific review as De Melker would **have** us *believe*. <BNC:FSY 1369>

(25)   This is direct experience, but it is not drama -- not until there is some pretence involved, some symbolic representation, some intention to **make** *believe*. <BNC:AM6 126>

Let us now turn to one of the only causative constructions whose collexemes have attracted some attention in the literature, [X *MAKE* Y V_{pp}], and examine Table 9, which provides a complete list of the collexemes distinctive for native English (upper part) and learner English (lower part). In native English, the construction is highly restricted in terms of lexical preferences. Only four verbs are distinctive, viz. *known*, *felt*, *recognised* and *swallowed*, and of the 19 tokens represented by these verbs, 12 correspond to an occurrence of *known*, as in (26). Two other verbs are found in BNCw with [X *MAKE* Y V_{pp}], although they are not distinctive for native English, namely *understood* and *heard*. It will be noticed that, with the exception of *swallowed*, all these verbs back up Van Ek & Robat's (1984: 327) claim that the construction is used with verbs "denoting the exercise and recognition of influence in the widest sense".

(26)   The exercise was regarded by the then minister for health in Scotland, Michael Forsyth, as a piece of "action" research, the results of which would be **made** *known* at intervals during the evaluation. <BNC:FT3 859>

Collexemes	Obs. freq. BNCw	Obs. freq. ICLE	Exp. freq. BNCw	Exp. freq. ICLE	Distinctiveness	Preferred variety
known	12	8	7.21	12.79	2.1085	BNCw
felt	5	1	2.16	3.84	1.7025	BNCw
recognised	1	0	0.36	0.64	0.4429	BNCw
swallowed	1	0	0.36	0.64	0.4429	BNCw
understood	2	9	3.97	7.03	0.8109	ICLE
heard	1	5	2.16	3.84	0.5423	ICLE
legalised	0	2	0.72	1.28	0.3926	ICLE
undone	0	2	0.72	1.28	0.3926	ICLE
acted	0	1	0.36	0.64	0.1943	ICLE
based	0	1	0.36	0.64	0.1943	ICLE
broken	0	1	0.36	0.64	0.1943	ICLE
brought about	0	1	0.36	0.64	0.1943	ICLE
carried out	0	1	0.36	0.64	0.1943	ICLE

encouraged	0	1	0.36	0.64	0.1943	ICLE
fallen	0	1	0.36	0.64	0.1943	ICLE
obliged	0	1	0.36	0.64	0.1943	ICLE
projected	0	1	0.36	0.64	0.1943	ICLE
respected	0	1	0.36	0.64	0.1943	ICLE
stuck	0	1	0.36	0.64	0.1943	ICLE
thought	0	1	0.36	0.64	0.1943	ICLE

Table 9. Distinctive collexemes of [X *MAKE* Y V_{pp}] in BNCw and ICLE

In learner English, by contrast, there is a greater variety of collexemes occurring in the non-finite verb slot of [X *MAKE* Y V_{pp}]: 16 distinctive collexemes, most of which occur only once in the data, and 18 different verbs, for a total of 39 tokens. Calculating the type/token ratio, we obtain a result of 0.4615 in ICLE, against 0.2727 only in BNCw, which means that the lexical variation of the construction is much higher in learner English than in native English. Learners seem unaware of the strong lexical preferences of [X *MAKE* Y V_{pp}]. They use a whole series of verbs which do not belong to the semantic class of "exercise and recognition of influence" and, consequently, would be unlikely to occur in native English, e.g. *undone*, *broken*, *fallen* or *stuck*. The use of these verbs results in infelicitous constructions, as illustrated by (27) and (28).

(27)   Often fundamentalists try to affect society and **make** their norms *legalised*. <ICLE-FI>

(28)   Usually it automatically **makes** your dreams *based* on your experience. <ICLE-CZ>

The analysis presented in this section has shown that, while phenomena of over- and underuse and syntactic errors do pose problems for learners who want to use periphrastic causative constructions in English, another major stumbling block, largely ignored in the literature, is learners' lack of awareness of the lexical preferences displayed by the different constructions, which may result in awkward and unidiomatic usage. The next section offers three possible explanations for the discrepancies found between native and learner collostructions, viz. lack of register awareness, transfer from the mother tongue and inadequacy of teaching materials.

5. Explanations for learners' lexical infelicities

5.1. Lack of register awareness

Learners' lack of register awareness has been underlined by several linguists. More precisely, it has been demonstrated that learner writing tends to exhibit characteristics of informal, spoken English (see e.g. Altenberg & Tapper 1998 on the use of adverbial connectors or Gilquin & Paquot 2008 on language functions). This tendency seems to be at work in collostructions too. Thus, learners use quite a few instances of the [X *GET* Y *done*] collostruction, as in (29). The collostruction does occur in written English, cf. (30), but it turns out to be much more common in a corpus of spoken English (BNCs),[16] e.g. (31). This situation is represented by Figure 1.

(29)   It is a good thing if we manage to **get** more *done* in the same time. <ICLE-SW>

(30)   While on his own home ground he knows how to do things, or how to **get** them *done*, beyond the boundaries of his home ground he knows, at best, only that he should comply with the rules without necessarily understanding why or precisely what all of the relevant rules are: he may simply go through the motions. <BNC:BMP 714>

(31)   Well you **get** it *done* on the insurance, your insurance covers that. <BNC:KBB 6967>
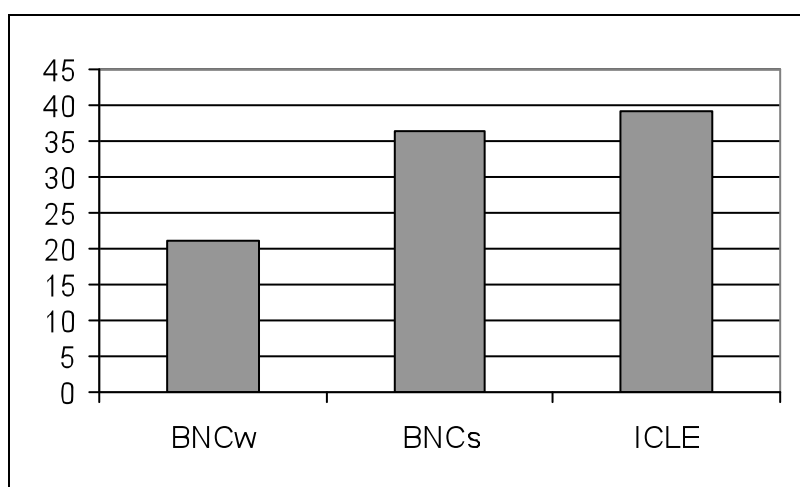
Figure 1. Proportion (%) of [X *GET* Y *done*] in native written English
(BNCw), native spoken English (BNCs) and learner English (ICLE)

The same tendency for learners to use collostructions more typical of speech than of writing is found with [X *MAKE* Y *feel*] which, as already mentioned, is distinctive for learner English. Table 10, however, shows that the distinctiveness only exists if one compares the data from ICLE with native written data (BNCw). In this case, the collexeme *feel* appears to be more frequent than expected in learner English (191 vs. 181.27), which explains the significant distinctiveness of the collostruction in ICLE. If, on the other hand, one compares the learner data with native spoken data (BNCs), *feel* turns out to be less frequent than expected in learner English (191 vs. 196.22), and the collostruction, more distinctive for native English – but not significantly so. In other words, while *feel* is significantly more distinctive for learner English when compared with native written English, there is no significant difference between learner English and native spoken English, which means that statistically the frequency of [X *MAKE* Y *feel*] in learner English is closer to its frequency in native speech than in native writing.

ICLE /	Obs. freq. BNC	Obs. freq. ICLE	Exp. freq. BNC	Exp. freq. ICLE	Distinc- tiveness	Preferred variety
BNCw	21	191	30.73	181.27	1.6205	ICLE
BNCs	121	191	115.78	196.22	0.5596	BNCs

Table 10. Distinctiveness of [X *MAKE* Y *feel*] in ICLE vs. BNCw and ICLE vs. BNCs

These two examples suggest that learners have difficulty distinguishing between collostructions that are likely to be found in an academic essay and those that are more likely to occur in a conversation. This could explain why there is so little overlap between the collexemes used by learners and native speakers in their writings. However, not all cases of discrepancy between ICLE and BNCw can be explained by means of a difference in register.

5.2. Transfer from the mother tongue

Another possible explanation for the lexical infelicities found in learner writing is transfer from the mother tongue. While in the present study no distinction has been made between the different learner populations, it was shown in Gilquin (2000/2001) that the overuse of dynamic, as opposed to stative, verbs in the non-finite verb slot of causative constructions

with *make* by French-speaking learners may be due to the influence of French, where the equivalent causative construction, [X *FAIRE* V_{inf} Y], usually contains a dynamic verb (cf. Table 11). As a result, the high frequency, in French-speaking learners' essays, of sentences like (32), which are comparatively rare in native writing, may be related to the presence of similar sentences in native French.

(32)   There are only two subjects on earth that have the power to pervert people and to **make** them *do* anything crazy, those two subjects are love and money. <ICLE-FR>

	US	ICLE	FR
Dynamic	41%	68%	71%
Stative	59%	31%	29%

Table 11. Dynamic and stative verbs in the non-finite verb slot of the *make/faire* causative construction in native American English (US), learner English (ICLE) and native French (FR) (based on Gilquin 2000/2001: 110)

5.3. Inadequacy of teaching materials

Finally, the inadequacy of teaching materials may also be blamed for learners' lack of awareness of the lexical preferences displayed by English periphrastic causative constructions. For one thing, as already noted, information about such preferences is conspicuously absent from the literature, let alone from the reference tools available to learners. Grammars, both pedagogical and scientific, tend to focus on syntactic issues, in particular complementation and passivisation (see Gilquin 2004), and textbooks often deal with periphrastic causative verbs in a section on verb patterns, together with other verbs that can be used in the same structure (*see, require, believe,* etc). For another thing, the examples that are used in teaching materials to illustrate the different causative constructions often sound artificial, using collexemes which are very unlikely to occur in authentic English. Examples (33) to (35), collected from various grammars, illustrate this tendency.

(33)   They **made** me *do* it. (Murphy 1985: 110)
(34)   She didn't **have** him *kill* the rat. (Cupers & Loriaux 1977: 26)

(35)   As soon as our guests arrived, we **had** them *peeling* potatoes and *slicing* beans.
        (Gusdorf & Lewis 2002: 536)

With no adequate resources to turn to and no native-speaker intuitions to rely on, no wonder learners are at a loss to make felicitous lexical choices when using causative constructions.

6. Conclusion

In this article, we have seen that English periphrastic causative constructions display strong lexical preferences for certain non-finite verbs, which, with rare exceptions, are not recorded in the literature, let alone in pedagogical grammars and textbooks. Foreign learners, with no native-speaker intuitions and no appropriate tools to rely on, tend to produce unidiomatic constructions, as the analysis of a learner corpus reveals. While such constructions are unlikely to hinder communication, they contribute to the foreign-soundingness of learners' production and should therefore be eradicated if one wants to attain native-like proficiency. This eradication implies at least two steps. The first step, which has been carried out here for causative constructions, involves establishing the lexical preferences found in native language and comparing them with learners' preferences, using corpus data and appropriate statistical methods. The second step would be to bring these findings (or at least, some of them) into the classroom. Howarth (1998: 30) notes that "teachers and materials writers are paying increasing attention to the necessity of learners to acquire knowledge of collocations and are aware that this component of competence should be addressed explicitly". However, it has been demonstrated that there is still room for substantial improvement (e.g. Biber et al. 2004, Meunier & Gouverneur 2007). Presenting learners with authentic and typical examples and drawing their attention to the most frequent and relevant collocations/collostructions is a sine qua non for "phraseologically improved" teaching materials. Until learners have such resources at their disposal, one can hardly blame them for being unaware of phenomena that partly escape the conscious attention of native speakers themselves.

   A study such as this one could also have implications on a more theoretical level. Thus, constructionists claim that during the acquisition of a second language, learners increasingly rely on constructions (cf. Goldberg 2006). One could wonder whether such a quantitative increase goes hand in hand with a qualitative improvement, namely the use of increasingly idiomatic constructions, closer in their lexical features to the preferences displayed by native speakers. A longitudinal corpus, representing the production of learners at

different developmental stages, combined with the method of distinctive collexeme analysis applied here, would make it possible to answer this question and hence go one step further in our understanding of the mechanisms underlying the acquisition of constructions and, more generally, of language. This, in turn, could inform more pedagogical decisions, thus resulting in mutual enrichment between theory and practice.

References

Adamczewski, Henri and Claude Delmas. 1982. *Grammaire linguistique de l'anglais*. Paris: Armand Colin.

Allerton, David J. 1984. "Three (or four) levels of word cooccurrence restriction". *Lingua* 63.1: 17-40.

Altenberg, Bengt and Sylviane Granger. 2001. "The grammatical and lexical patterning of MAKE in native and non-native student writing". *Applied Linguistics* 22.2: 173-194.

Altenberg, Bengt and Marie Tapper. 1998. "The use of adverbial connectors in advanced Swedish learners' written English". In Sylviane Granger (ed.) *Learner English on Computer* (pp. 80-93). London & New York: Addison Wesley Longman.

Biber, Douglas, Susan Conrad and Viviana Cortes. 2004. "*If you look at…*: Lexical bundles in university teaching and textbooks". *Applied Linguistics* 25.3: 371-405.

Burnard, Lou. 2000. *Reference Guide for the British National Corpus (World Edition)*. Available at http://www.natcorp.ox.ac.uk/docs/userManual (last accessed on February 1, 2007).

Cupers, Jean-Louis and Christian Loriaux. 1977. *Grammaire fondamentale de l'anglais. Abrégé des faits généraux de la langue anglaise*. Bruxelles: De Boeck.

Gilquin, Gaëtanelle. 2000/2001. "The Integrated Contrastive Model. Spicing up your data". *Languages in Contrast* 3.1: 95-123.

Gilquin, Gaëtanelle. 2004. *Corpus-based cognitive study of the main English causative verbs. A syntactic, semantic, lexical and stylistic approach*. Unpublished PhD Thesis. Louvain-la-Neuve: Université catholique de Louvain.

Gilquin, Gaëtanelle. 2006. "The verb slot in causative constructions. Finding the best fit". *Constructions* SV1-3/2006 (www.constructions-online.de, urn:nbn:de:0009-4-6741, ISSN 1860-2010).

Gilquin, Gaëtanelle. In preparation. "Making sense of collostructional analysis. On the interplay between verb senses and collexemes".

Gilquin, Gaëtanelle and Magali Paquot. 2008. "Too chatty: learner academic writing and register variation". *English Text Construction* 1.1: 41-61.

Goldberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.

Granger, Sylviane, Estelle Dagneaux and Fanny Meunier (eds). 2002. *The International Corpus of Learner English. Handbook and CD-ROM.* Version 1. Louvain-la-Neuve: Presses universitaires de Louvain. Available from http://www.i6doc.com.

Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot (eds). Forthcoming. *International Corpus of Learner English. Handbook and CD-ROM.* Version 2. Louvain-la-Neuve: Presses universitaires de Louvain.

Gries, Stefan Th. 2004. *Coll.analysis 3. A program for R for Windows 2.x.* Available at http://www.linguistics.ucsb.edu/faculty/stgries (last accessed on February 1, 2007).

Gries, Stefan Th. and Anatol Stefanowitsch. 2004a. "Extending collostructional analysis. A corpus-based perspective on 'alternations'". *International Journal of Corpus Linguistics* 9.1: 97-129.

Gries, Stefan Th. and Anatol Stefanowitsch. 2004b. "Covarying collexemes in the *into*-causative". In Michel Achard and Suzanne Kemmer (eds) *Language, Culture, and Mind* (pp. 225-236). Stanford, CA: CSLI.

Gusdorf, Florent and Stephen Lewis. 2002. *Grammaire anglaise.* Palaiseau: Editions de l'Ecole polytechnique.

Halliday, Michael A. K. 2004. *An Introduction to Functional Grammar*. Third Edition (revised by C.M.I.M. Matthiessen). London: Edward Arnold.

Howarth, Peter 1998. "Phraseology and second language proficiency". *Applied Linguistics* 19.1: 24-44.

Lehmann, Hans-Martin, Sebastian Hoffmann and Peter Schneider. 2002. *BNCweb: A web-based interface to the British National Corpus. Version 2.0.* Zurich: Zurich University. See http://homepage.mac.com/bncweb/home.html (last accessed on February 1, 2007).

Liu, Eric T.K. and Philip M. Shaw. 2001. "Investigating learner vocabulary: a possible approach to looking at EFL/ESL learners' qualitative knowledge of the word". *International Review of Applied Linguistics in Language Teaching* 39.3: 171-194.

Meunier, Fanny and Céline Gouverneur. 2007. "The treatment of phraseology in ELT textbooks". In Encarnación Hidalgo, Luis Quereda and Juan Santana (eds) *Corpora in the Foreign Language Classroom. Selected Papers from the Sixth International Conference on Teaching and Language Corpora (TaLC6), University of Granada, Spain, 4-7 July, 2004* (pp. 119-139). Amsterdam & New York: Rodopi.

Murphy, Raymond. 1985. *English Grammar in Use. A Self-study Reference and Practice Book for Intermediate Students*. Cambridge: Cambridge University Press.

Scott, Mike. 1999. *WordSmith Tools 3*. Oxford: Oxford University Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. "Collostructions: investigating the interaction of words and constructions". *International Journal of Corpus Linguistics* 8.2: 209-243.

Stocker, Markus. 1990. *The Causative in Middle and Modern English*. Unpublished PhD Thesis. Muttenz: Universität Basel.

Stubbs, Michael. 1993. "British traditions in text analysis. From Firth to Sinclair". In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds) *Text and Technology. In Honour of John Sinclair* (pp. 1-33). Philadelphia & Amsterdam: John Benjamins Publishing Company.

Tomasello, Michael. 2003. *Constructing a Language. A Usage-based Theory of Language Acquisition.* Cambridge, Massachusetts & London, England: Harvard University Press.

Turton, Nigel D. and John B. Heaton. 1996. *Longman Dictionary of Common Errors*. London: Longman.

Van Ek, Jan A. and Nico J. Robat. 1984. *The Student's Grammar of English*. Glasgow: Bell & Bain Ltd.

Wong, Sau-Ling Cynthia. 1983. "Overproduction, underlexicalisation and unidiomatic usage in the *make* causatives of Chinese speakers: a case for flexibility in interlanguage analysis". *Language Learning and Communication* 2.2: 151-165.