

Light verb constructions in spoken L2 English: An exploratory cross-sectional study

Gaëtanelle Gilquin
Université catholique de Louvain

Abstract

Light verb constructions (LVCs), that is, combinations like *take a walk* or *make a choice*, are often claimed to be problematic for non-native speakers of English. In this paper, spoken data from the Trinity Lancaster Corpus are used to explore the use of these constructions across different sections of the corpus, representing different proficiency levels (from lower intermediate to upper advanced) as well as different types of acquisitional contexts, namely English as a foreign language (EFL) and English as a second language (ESL). The results of the study reveal a tendency towards an increased frequency of LVCs, as well as more complex and abstract uses, as we move from an intermediate to an advanced level and from an EFL to an ESL context. For the EFL speakers, this development corresponds to a better approximation to native English. For the ESL speakers, on the other hand, LVCs seem to have become 'constructional teddy bears', being used more often than by the native speakers themselves.

Keywords: light verb construction, speech, EFL, ESL, proficiency level, developmental patterns

1. Introduction

High-frequency verbs like *GET*, *GIVE*, *MAKE* or *TAKE* are often "lexical teddy bears" (Hasselgren 1994) for non-native (L2) users of English, being part of their repertoire of preferred words because they are considered to be safe and easy to use. Yet, there is at least one use of high-frequency verbs that tends to be more problematic in L2 English, namely their use in light verb constructions (LVCs), which consist in the combination of a light verb (typically a high-frequency verb) and an eventive noun (e.g. *have a look*, *give a smile*, *take a walk*). Several corpus-based studies of L2 English have brought to light both quantitative problems (underuse or overuse) and qualitative problems (misuse) related to the use of these constructions. In this paper, I would like to explore aspects of LVCs in L2 English that have been under-researched so far. First, I will look at the way LVCs evolve as proficiency increases, from a lower intermediate to an upper advanced level. Second, I will compare LVCs in English as a Foreign Language (EFL) and English as a Second Language (ESL). These two aspects will be investigated in L2 speech thanks to a new corpus resource, the Trinity Lancaster Corpus (TLC). The focus will be on the frequency of LVCs, although some more formal aspects will also be briefly examined.

The paper is organized as follows. Section 2 provides some background information on light verb constructions in English, with a focus on the literature written from the perspective of learner corpus research. The next section describes the corpus data that were used and the methodology that was applied. The results of the analyses are found in Sections 4 and 5. The article ends with a discussion (Section 6) and some concluding remarks (Section 7).

2. Light verb constructions: definition and use in L2 English

Light verb constructions, also called ‘support verb constructions’ (Langer 2004), ‘stretched verb constructions’ (Allerton 2002) or ‘complex predicates’ (Traugott 1999), are normally made up of two elements. The first one is a ‘delexical’ verb that, in the context of the construction, is semantically light, i.e. contributes very little meaning to the construction.¹ The second element is an eventive, ‘deverbal’ noun, which is related to a verb and carries the main semantic weight of the construction. While most definitions of LVCs agree on the presence of these two elements,² they differ, sometimes widely, in their descriptions of the other characteristics of the construction. Thus, next to high-frequency and general verbs, some definitions cover more specific verbs as well (e.g. *commit* as in *commit a murder* in Langer 2004). As for the deverbal noun, its relation to a verb can take several forms: the correspondence may be identical, when the noun has the same form as a verb (cf. above examples: *have a look*, *give a smile*, *take a walk*); it can be derived from a verb, as in *make a choice* or *take a decision*; or it may be semantically, but not formally, related to a verb, e.g. *make an effort*, *have a game* (cited in Quirk et al. 1985: 751). Not all definitions of LVCs accept the three types of nouns (especially the last one may not be covered by the definition). Similarly, there may be differences in whether or not the definition includes nouns other than those preceded by an indefinite article (e.g. *take the lead*, *do some homework*, *make use*) and nouns preceded by a preposition (e.g. *take into consideration*). Finally, the construction may have a passive meaning, especially with the verb *have*, cf. *have a fright*, corresponding to ‘be frightened’ rather than to ‘frighten’ (Quirk et al. 1985: 751). In this paper, I will include under the label ‘light verb constructions’ all constructions made up of a high-frequency verb (among *MAKE*, *TAKE* and *GIVE*)³ and a noun phrase including a deverbal noun (formally identical to a verb, derived from one or semantically related to one; preceded by a determiner or not) which express an active meaning. This definition thus excludes constructions in which the deverbal noun is part of a prepositional phrase (like *take into consideration*) and those that have a passive meaning (like *have a fright*).

Several corpus-based studies have investigated LVCs in L2 English.⁴ While several aspects of these constructions have been dealt with, including their non-standard uses (e.g. Nesselhauf 2005 on EFL, Scheepers 2017 on ESL), here, in accordance with the focus of this paper and for the sake of brevity, I will only review those studies that have considered the frequency of LVCs. The findings of these studies tend to diverge. In some of them non-native speakers appear to underuse LVCs in comparison to native speakers (e.g. Altenberg & Granger 2001, Shirato & Stapleton 2007), whereas in others it is a pattern of overuse that is brought to light (e.g. Neff-van Aertselaer & Bunce 2008, Wang 2016). These diverging results could be explained by the fact that learner populations with different mother tongues (L1) are examined, which, in case of transfer from the L1 (see Osborne 2015 on transfer), could lead to differences in the use of LVCs (Altenberg & Granger 2001 investigate Swedish and French-speaking learners, while Neff-van Aertselaer & Bunce 2008 investigate Spanish learners; see also Hoffmann et al. 2011 on the difference in frequency between several ESL varieties). That this cannot be the whole story, however, appears from a comparison of the results obtained by

¹ Whether ‘light verbs’ are indeed ‘light’ in the sense of contributing no or little meaning is a matter of debate which I will not enter here. See, e.g., Brugman (2001) or Butt & Geuder (2001) on the semantic contribution of light verbs.

² Brugman’s (2001) definition also includes cases where the noun is ‘deadjectival’ (e.g. *have the intelligence*, where the noun is related to the adjective ‘intelligent’).

³ Other possible light verbs like *DO* or *HAVE* were not included in the analysis because of the huge amount of manual editing that this would have required (cf. auxiliary uses of these verbs).

⁴ For studies approaching LVCs from a Second Language Acquisition (not necessarily corpus-based) perspective, see e.g. Wang (2011), Szudarski (2012), Kittigosisin & Phoocharoensil (2015) or Szudarski & Carter (2016).

Altenberg & Granger (2001) and Wang (2016) for Swedish learners: while the former found an underuse of LVCs, the latter found an overuse. The verbs that are considered in the analysis may also have an impact on the results. This is suggested by Juknevičienė (2008), who shows that, of the five verbs she investigates (*HAVE*, *DO*, *MAKE*, *TAKE*, *GIVE*), only two (*MAKE* and *GIVE*) display a significant underuse among Lithuanian learners, the results for the other three verbs being statistically non-significant. This could possibly explain the difference between Altenberg & Granger (2001), who focused on the verb *MAKE*, and Neff-van Aertselaer & Bunce (2008), who focused on *have*. Other factors that could account for the divergent results of these studies include: differences in proficiency levels, in the native corpus used as a reference, in the genre/register under study and in the definition of the LVC. Many other factors could explain these results, for, as Ellis (1994: 49) puts it, “the factors that can bring about variation in learner output are numerous, perhaps infinite”. More generally, the variety of results obtained by these studies may be seen as a reflection of the paradox highlighted by Altenberg & Granger (2001), namely that high-frequency verbs are likely to be overused by non-native speakers, because they represent core words that seem safe to use (cf. the ‘lexical teddy bear’ phenomenon described in Section 1), but at the same time they run the risk of being avoided because non-native speakers may be unsure about some of their idiomatic uses (cf. Sinclair 1991: 79).

Most of the studies on LVCs in L2 English referred to above are based on *written learner* corpora and investigate *EFL learners* at an *advanced* level of proficiency. By contrast, this paper explores the use of LVCs in *spoken* English as a *foreign and second language* from a *developmental* point of view. Studies on the use of LVCs in L2 speech are rare in the literature. Shirato & Stapleton (2007) is an example of such a study, but delexical verbs are only one of the features they examine and their notion of delexical verb is much wider than the one used here (for example, it includes phrasal verbs and *get*-passives). The study of LVCs in L2 speech will be carried out from a twofold perspective. The first one will examine the evolution of LVCs across proficiency levels, from a lower intermediate to an upper advanced level, corresponding to a B1 level to a C2 level in the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001). Such a developmental approach to LVCs can be found in Laufer & Waldman (2011), who compare three groups of Hebrew-speaking learners representing different proficiency levels – basic, intermediate and advanced. They used written data, however, not spoken data. The second perspective will compare the use of LVCs in EFL and ESL situations, thus considering the possible impact of the context of acquisition. Quite a few studies over the last few years have looked into the links between EFL and ESL (in the sense of institutionalized second-language varieties of English, also called ‘New Englishes’), e.g. Mukherjee & Hundt (2011) or Gilquin (2015). Most of these studies, however, have used written corpora – partly, as will be explained in Section 3, for lack of adequate resources to compare EFL and ESL speech. Nesselhauf (2009), for example, one of the earliest corpus-based comparisons of EFL and ESL, mainly relies on written data (with some additional ESL spoken data for low-frequency phenomena) to investigate a number of phraseological units, among which a few LVCs (e.g. *take into consideration*, *have the intention*).

Both of these perspectives – developmental and EFL/ESL comparison – will help us get a better understanding of how LVCs are used in L2 spoken English. The hypothesis that will be tested is that some improvement in the use of LVCs should be visible as proficiency increases and as we move from EFL to ESL. The potential link between increase in proficiency level and improvement in the use of LVCs does not require a great deal of explanation. Although, as we know, second language acquisition is not necessarily linear (see Larsen-Freeman 1997), we can expect at least some degree of improvement as we move from, say, B1 to C2 (cf. Bulté & Housen 2015: 50). As for the potential link with the acquisitional context

(EFL/ESL), it relies on the amount and type of input received: while in EFL contexts input is mainly limited to ‘instructional English’, i.e. English as used in the classroom, in ESL contexts input can also be found in everyday life, since English is an official language used in the media, administration, education, etc. The increased and more authentic input available in ESL contexts can be hypothesized to lead to a better knowledge of the language (see Gilquin 2016 and the references therein), and hence of LVCs. In this study, improvement will mainly be measured in terms of frequency, with frequency in English as a native language (ENL) being considered the target to reach. Some formal aspects of LVCs will also be briefly considered in an attempt to check whether improvement can occur at other levels as well. These aspects will include the use of concrete vs abstract deverbal nouns, since more abstract (i.e. less imageable) words are arguably indicative of greater lexical proficiency (see Crossley et al. 2011). We will also examine passivization of the light verb, as the use of passive forms is a common measure of grammatical complexity (cf. Bulté & Housen 2012), as well as premodification of the deverbal noun, since “[a] simple noun phrase has a determiner and head noun” and “[a]dditions or modifications to these patterns result in complex grammar” (Biber et al. 2011: 12). These features of (lexical and syntactic) complexity will make it possible to combine the quantitative analysis based on the frequency of LVCs with a more qualitative touch.

3. Data and methodology

The Trinity Lancaster Corpus (TLC) is a new resource which is made up of transcriptions of the Graded Examination in Spoken English organized by Trinity College London (see Gablasova et al. this volume). It includes data produced by L2 speakers from different linguistic and cultural backgrounds and different L2 proficiencies, who participated in various speaking tasks. The TLC Sample consists in six components produced by Chinese, Indian, Italian, Mexican, Russian and Spanish speakers with proficiency levels ranging from the threshold level B1 to the mastery level C2. Components with less than 200,000 words of L2 production were excluded for the purposes of this study because of their small sizes. This left three EFL components – Italian (TLC-ITA), Mexican (TLC-MEX) and Spanish (TLC-SPA) – and one ESL component – Indian (TLC-IND) – for analysis. This selection shows limitations that should be kept in mind when considering the results of the study. The EFL data include learners who all have a Romance L1 background (Italian or Spanish), and the results might therefore not be representative of EFL learners in general. As for the ESL data, they correspond to one variety only, namely Indian English, which, again, might not be representative of ESL as a whole. However, the selected data do correspond to an EFL and ESL context, respectively, as confirmed by the background questionnaires filled in by the candidates. The metadata accompanying the corpus indicate that the candidates lived in their home country (Italy, Mexico, Spain or India) and displayed a good gender distribution (with slightly more female than male candidates).

The cross-sectional nature of the TLC, gathering data “from different categories of learners at a single point in time” (Granger 2008: 263), not only makes it possible to compare EFL and ESL learners, but also to adopt a quasi-longitudinal design, which looks into the evolution of learner language on the basis of different learners corresponding to different proficiency levels (unlike a longitudinal design, which compares the same learners at different levels; see Meunier 2015 on the distinction between longitudinal and quasi-longitudinal designs in learner corpus research). In the subsample used here, the full range of proficiencies available in the TLC was taken into account, from B1 to C2, but the C1 and C2 levels were

grouped together because of the relatively limited number of C1 and C2 candidates.⁵ Only those speaking tasks that were found at all levels, i.e. discussions and conversations, were kept for analysis.

The EFL and ESL analyses were based on the utterances produced by the candidates. As for the utterances produced by the examiners (in the discussions and conversations of the whole TLC Sample), they were used as a reference corpus against which to compare the L2 data. This reference corpus (TLC-EXA) represents native English, since all the examiners were native speakers of English, and it has the advantage of being comparable to the L2 corpus in terms of genre and tasks. It should be emphasized that some examiners took part in several exams. For the purposes of statistical analysis, however, all cases of examiner speech were treated as independent events. While this may to some extent compromise the independence of observations if the observation is construed as a unique speaker, the rationale for this type of treatment lies in the fact that each examiner-candidate interaction can be considered a unique and independent speech event.

Table 1 provides information about the number of speakers and word counts of the different sets of data that were used in this study. The reference corpus amounts to almost 800,000 words and the L2 corpus to almost one million words, of which over 700,000 represent EFL and the remainder corresponds to ESL. The distribution according to proficiency level is also shown in the table.

Table 1. Composition of the TLC subsample per variety and per proficiency level: number of speakers and word count (including mean number of words per speaker and standard deviation)

	Variety	ENL	EFL			ESL
Level		TLC-EXA	TLC-ITA	TLC-MEX	TLC-SPA	TLC-IND
B1	No. of speakers	NA	100	100	100	116
	No. of words		79,781	81,488	74,256	102,668
	- Mean		797.81	814.88	742.56	885.07
	- SD		140.64	201.53	150.70	277.00
B2	No. of speakers	NA	100	70	100	107
	No. of words		102,390	71,648	94,702	117,019
	- Mean		1023.90	1023.54	947.02	1093.64
	- SD		196.51	225.68	196.09	265.38
C1/C2	No. of speakers	NA	60	60	60	37
	No. of words		75,415	71,861	64,672	50,074
	- Mean		1256.92	1197.68	1077.87	1353.35
	- SD		259.39	225.04	257.73	293.43
Total	No. of speakers	1,244 (93)⁶	260	230	260	260
	No. of words	785,002	257,586	224,997	233,630	269,761
	- Mean	631.03	990.72	978.25	898.58	1037.54
	- SD	202.36	262.41	266.18	237.19	317.52

As appears from the table, the structure of the TLC, and of the selected subsample in particular, allows for different types of comparison to be drawn: between native and non-native

⁵ In the examples, however, information about the exact level (C1 or C2) will be provided through the reference code (which also indicates the TLC component from which the example is taken – TLC-EXA, TLC-IND, TLC-ITA, TLC-MEX or TLC-SPA).

⁶ 93 is the number of unique examiners. However, as explained above, for statistical purposes all cases of examiner speech (1,244) were counted as individual events.

English, between EFL and ESL, between Italian, Mexican and Spanish EFL learners, as well as between three proficiency levels. The comparison between EFL and ESL is especially interesting, since studies comparing EFL and ESL have so far mainly analysed written data, for the practical reason that truly comparable (sub)corpora of EFL and ESL are only available for writing. For speech, the few comparisons of EFL and ESL have had to rely on data that present different characteristics, thus potentially influencing the results (see Gilquin 2015). The TLC can therefore be said to offer an unprecedented opportunity to study EFL and ESL interactive speech on the basis of truly comparable corpus data.⁷

I searched the corpus by means of the web-based program Sketch Engine (Kilgarriff et al. 2014). After selecting the relevant variables (speaker, candidate proficiency, country and task), I extracted all the occurrences of three lemmas, namely *MAKE*, *TAKE* and *GIVE*. Each of the c. 6,500 concordance lines thus extracted was examined manually to determine whether the verb was used in a LVC or not, applying the definition outlined in Section 2.⁸ In this stage of disambiguation, context was crucial. Both (1) and (2), for example, include the combination *make games*. In the first sentence, the meaning is ‘create/produce games’, as appears from the reference to a game developer. In the second sentence, on the other hand, *make games* has the sense of ‘playing’, a verbal meaning that is expressed by the noun ‘game’. The second example thus corresponds to a LVC, but not the first one. As illustrated by the *make games* example, non-standard constructions were kept for analysis, as long as it was sufficiently clear that they met the criteria for identifying LVCs. Occasional occurrences of a high-frequency verb with a deverbal noun in the candidate’s L1 were also taken into account, cf. *make the ofrendas*, where *ofrendas* is the Spanish word for ‘offerings’. In total, 1,503 instances of LVCs were identified. Table 2 shows the distribution of these constructions per variety and per proficiency level.

- (1) *I’m going to trying gaming I’m want to **make games** I want to be a game developer right* (TLC-IND-B2)
 (2) *and we make a we **make games** with the children* (TLC-MEX-B1)

Table 2. Distribution of the light verb constructions in the TLC subsample per variety and per proficiency level

Variety Level	ENL	EFL			ESL
	TLC-EXA	TLC-ITA	TLC-MEX	TLC-SPA	TLC-IND
B1	NA	34	63	34	99
B2	NA	53	62	81	114
C1/C2	NA	69	66	61	91
Total	676	156	191	176	304

⁷ The International Corpus Network of Asian Learners of English (ICNALE) is made up of both EFL and ESL data, and includes a spoken component. However, this component consists in monologic speech. In addition, as its name indicates, only L2 speakers of Asian origin contributed data to this corpus.

⁸ Because of the large number of hits, double coding of the data was not an option. However, the criteria described in Section 2 were strictly adhered to. In case of doubt, a tentative decision was made and the concordance line was flagged with a question mark. After having examined all the concordance lines, a final decision was made for the pending cases, based on the coding of the other concordance lines and with consistency as a guiding principle.

4. Light verb constructions in EFL

The first part of this study focuses on EFL learners and how their use of LVCs may differ from that of native speakers and vary according to proficiency. The EFL learners are examined as a whole group, but the statistics provided (Wilcoxon rank-sum test or Kruskal-Wallis)⁹ take individual differences into account. For some of the analyses, the three verbs under study (*MAKE*, *TAKE* and *GIVE*) are distinguished, but otherwise LVCs are considered as a category. Sections 4.1 and 4.2 focus on quantitative aspects (frequency), while Section 4.3 deals with more qualitative aspects (semantic abstraction, syntactic complexity).

4.1. Frequency of light verb constructions in EFL

Figure 1 is a boxplot showing the relative frequency of LVCs per speaker in the EFL and ENL data and including all data points represented as a cloud. It will be used as an illustration of some of the features of the data examined here. The first observation is that many of the speakers do not produce any LVC at all (cf. all the dots with a relative frequency equal to zero). This is the case for 53% of the Mexican candidates, 63% of the Spanish candidates and 65% of the Italian candidates.¹⁰ Among those candidates that do produce LVCs, the frequency of the construction tends to be very low. As a result, the median (*Mdn*) often equals 0 and will therefore not be reported, unless it is different from 0. The interquartile range (*IQR*), which includes the middle 50 percent of the data (represented by the box in the boxplot), amounts to 1.10 in EFL and 1.53 in ENL. As appears from the graph, the *IQR* actually corresponds to the third quartile (*Q3*), i.e. the lower 75% of the data, since the boxes start at 0. This is a characteristic of almost all boxplots that will be presented here, which is why I will report the *Q3* value rather than the *IQR* when describing differences between groups of speakers. The graph in Figure 1 also reveals the presence of many outliers, namely speakers who present an unusually high frequency of LVCs. The data are thus not normally distributed, as confirmed by a Shapiro-Wilk normality test ($W=.58$ for EFL and $.57$ for ENL, $p<.001$). As regards the comparison between EFL and ENL, it was explained in Section 2 that the results of studies diverge considerably regarding the overuse or underuse of LVCs in EFL (written) production. In the spoken data from the TLC subsample, it appears that overall, i.e. for the three light verbs taken together, there is little difference between the frequency of LVCs in EFL and that in ENL. In fact, according to the Wilcoxon rank-sum test, the difference is not statistically significant ($W=461750$, $p=0.66$).

⁹ These statistics were preferred to t-test and ANOVA because the data are not normally distributed (see Section 4.1).

¹⁰ The percentage for the native examiners cannot be determined, for lack of information about the individual examiners.

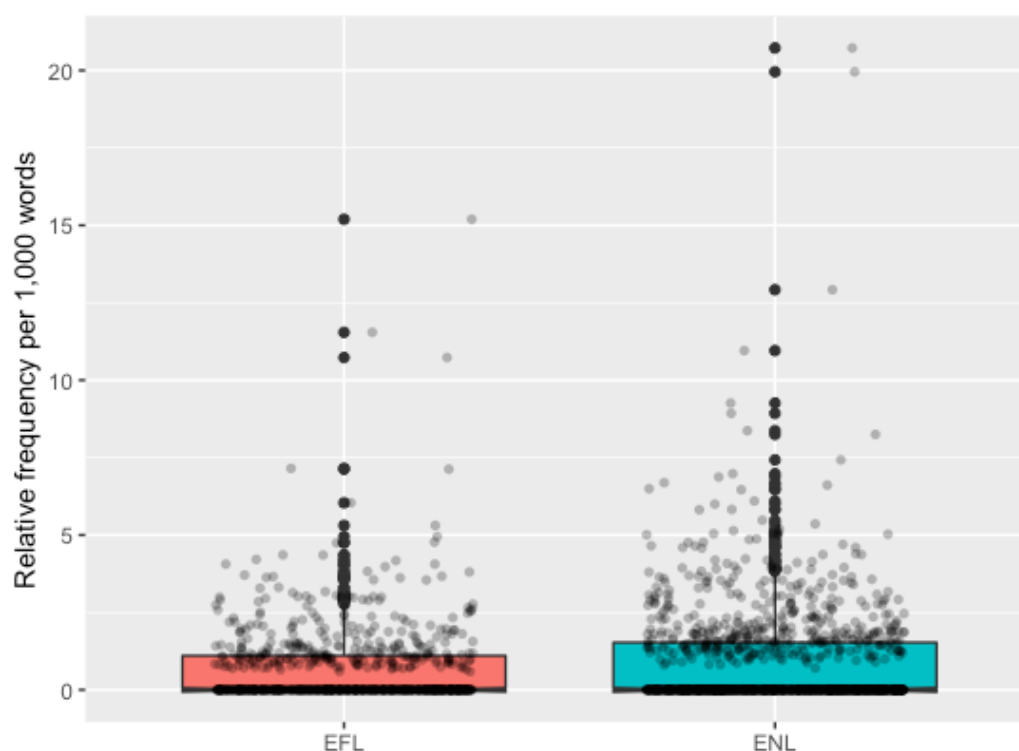


Figure 1. Relative frequency per 1,000 words of light verb constructions (all verbs) in EFL and ENL (all data points represented)

These general results will now be compared to the results per verb and per proficiency level. The next graphs will be simplified, in that, unlike Figure 1, they will not include the cloud of all data points. Their description will also be limited to the main findings. All boxplots will be notched, with the notch representing the 95% confidence interval of the median.

While we saw above that overall the frequency of LVCs is not statistically different in EFL and ENL, it appears from Figure 2 that, if we distinguish between the three verbs, the differences become significant according to the Wilcoxon rank-sum test.¹¹ In the case of *GIVE*, the difference between EFL ($Q3=0$) and ENL ($Q3=1.46$) is highly significant, with a small effect size ($p<.001$, $r=-.29$). In comparison with the native speakers, the learners can therefore be said to underuse LVCs with *GIVE*. With *MAKE*, the difference between EFL ($Q3=1.11$, $Mdn=0.60$) and ENL ($Q3=1.73$, $Mdn=1.00$) is also highly significant with a small effect size ($p<.001$, $r=-.17$), corresponding, again, to an underuse among the learners. While, in the case of *TAKE*, the difference between EFL ($Q3=1.14$) and ENL ($Q3=0.91$) is still highly significant with a small effect size ($p<.001$, $r=-.24$), this time it indicates an overuse among the learners. These results, pointing to an underuse of LVCs with *GIVE* and *MAKE*, and an overuse of LVCs with *TAKE*, are in line with a study like Juknevičienė (2008), showing that not all verbs behave in the same way. In fact, our results for speech coincide quite well with those obtained for writing by Juknevičienė (2008), who also found an underuse of LVCs with *MAKE* and *GIVE*, but not with the other verbs she examined (LVCs with *TAKE* displayed no significant difference, however, unlike the present study).

¹¹ Given the many speakers who did not produce a single occurrence of LVC (see above), only those speakers producing at least one LVC were considered in this calculation. P-values were adjusted using Bonferroni correction for multiple comparisons.

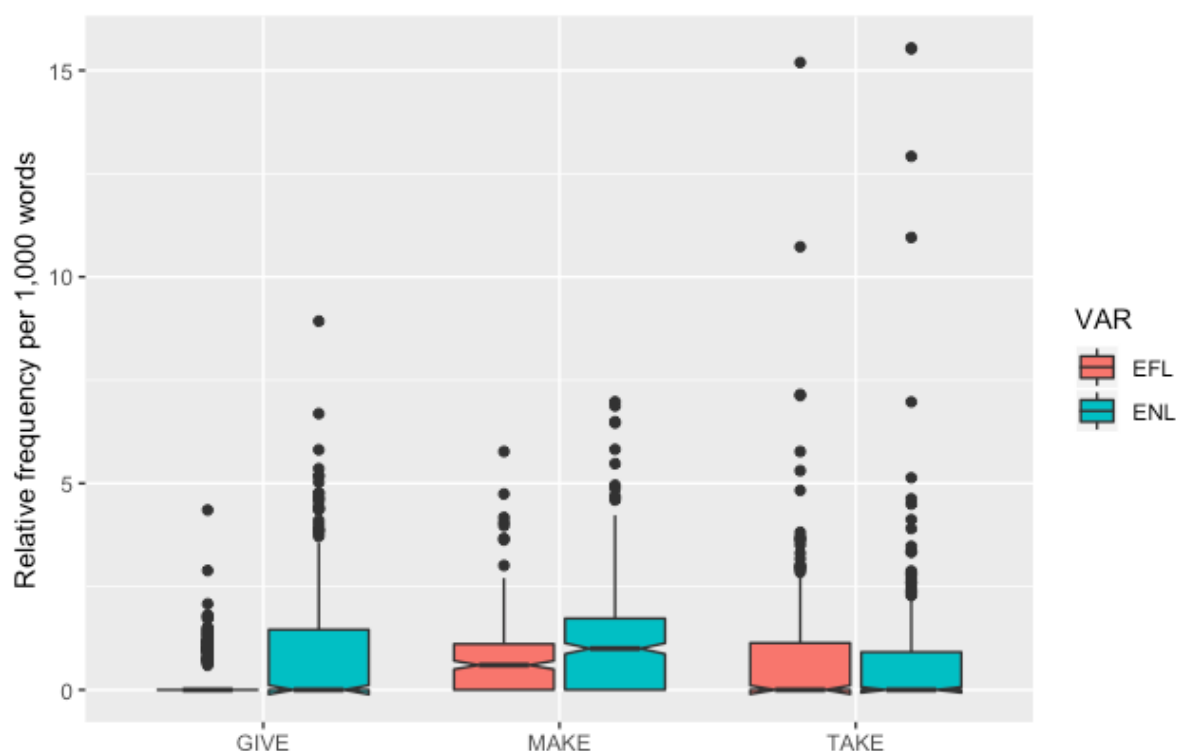


Figure 2. Relative frequency per 1,000 words of light verb constructions with *GIVE*, *MAKE* and *TAKE* in EFL and ENL

4.2. Frequency of light verb constructions across proficiency levels in EFL

Using the TLC as a quasi-longitudinal corpus, we can examine the evolution of the frequency of LVCs across groups of different proficiency levels. Taking the three verbs together, we observe a modest but steady rise in the relative frequency of LVCs as the proficiency level increases (see Figure 3), from a *Q3* of 0.99 at B1 level, through 1.09 at B2 level, to 1.46 (*Mdn*=0.74) at C level. This is to be compared with a *Q3* of 1.53 in native speech, a value that learners get closer to, the more proficient they become. The Kruskal-Wallis test reveals an overall highly significant difference in frequency between the CEFR groups ($H(3)=27.167$, $p<.001$), which a post hoc test¹² shows is situated between B1 and C ($B1<C$), between B1 and ENL ($B1<ENL$), and between C and ENL ($C<ENL$). While the hypothesis according to which an increase in proficiency corresponds to a more target-like frequency of LVCs in EFL is therefore confirmed, it also appears that there is still a quantitative gap between learners' and native speakers' use of LVCs (at least for learners at the B1 and C levels). This gap is still visible when a distinction is made between the three verbs. The Kruskal-Wallis post hoc test yields a significant difference between B1/C and ENL for all three verbs and, in addition, a significant difference between B2 and ENL for LVCs with *GIVE*. Note, however, that in the case of LVCs with *TAKE*, the frequency in native speech is lower than that in any EFL proficiency level, thus reflecting the overuse of LVCs with *TAKE* described in Section 4.1.

¹² Post hoc tests were performed for the Kruskal-Wallis test in order to see where exactly the significant differences lie, thanks to pairwise comparisons between the different groups (see Field et al. 2012: 681-684).

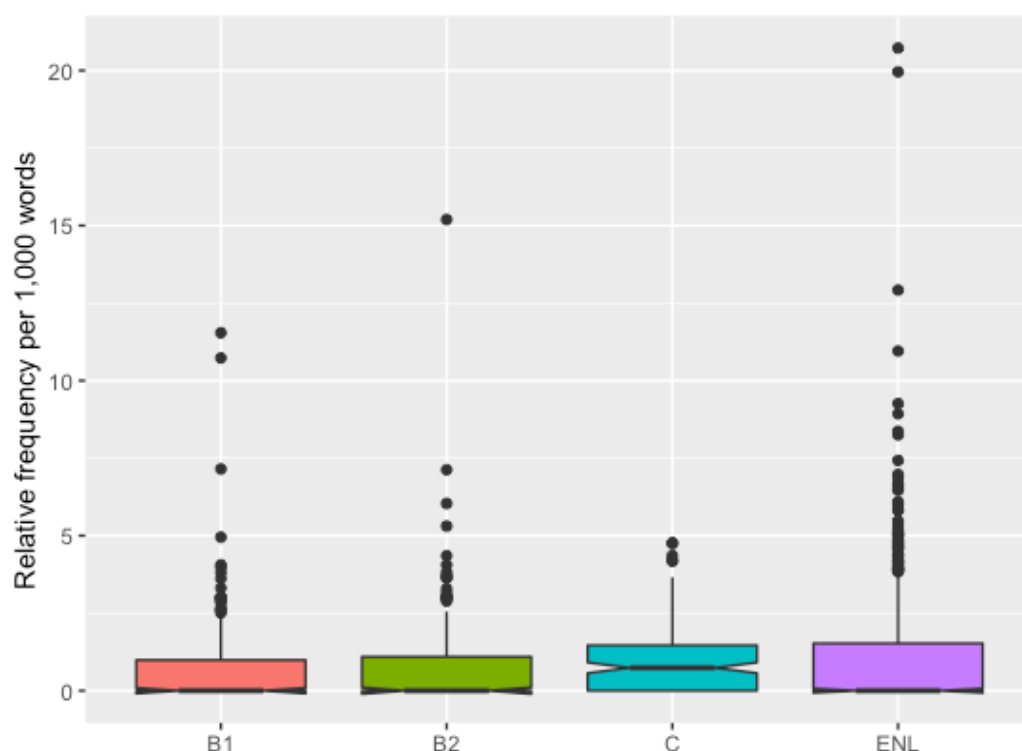


Figure 3. Relative frequency per 1,000 words of light verb constructions (all verbs) across EFL proficiency levels (B1, B2 and C) and in ENL

4.3. Formal aspects of light verb constructions in EFL

While a complete qualitative analysis of LVCs is beyond the scope of this paper, I would like to illustrate the fact that not only the frequency of the constructions may vary from one proficiency to the next, but also their form. More particularly, it appears that an increase in proficiency often goes hand in hand with an increase in the semantic abstraction and syntactic complexity of LVCs.

Although LVCs are, by definition, relatively abstract, consisting in the combination of a light verb (with no or little semantic content and largely arbitrary) and a deverbal noun (expressing a state or action in the form of a noun), some of the nouns used in LVCs can be more concrete than others. For example, the noun in *take a picture* can be said to be more concrete than that in *take a look*, which, arguably, is itself more concrete than that in *take a chance*. There seems to be a certain tendency among EFL learners to use more concrete deverbal nouns at lower proficiency levels. This is illustrated in Table 3 for LVCs with *MAKE* in the Spanish TLC subsample. The two nouns found at B1 level, *film* and *picture*, are relatively concrete. At B2 level, we see a mixture of more concrete (*film*) and more abstract nouns (*decision*, *campaign* and *effort*). At C level, finally, all deverbal nouns are rather abstract nouns: *change*, *discovery*, *conversation*, *difference*, *effort* and *mistake*.

Table 3. Nouns in light verb constructions with *MAKE* at different proficiency levels in the Spanish TLC subsample (with absolute frequencies)

B1	B2	C
<i>film</i> (4)	<i>decision</i> (6)	<i>change</i> (3)
<i>picture</i> (4)	<i>campaign</i> (2)	<i>discovery</i> (3)
	<i>effort</i> (2)	<i>conversation</i> (2)
	<i>film</i> (2)	<i>difference</i> (2)
		<i>effort</i> (2)
		<i>mistake</i> (2)

From a syntactic point of view, one way in which LVCs can become more complex across proficiency levels is through the use of the passive voice. Compare example (3) with examples (4) to (6). While in the first example the learner uses an active construction, in the other examples the construction is in the passive voice. In (4) it is the deverbal noun (direct object) that has become the subject of the passive construction. In (5), on the other hand, it is the indirect object that has become the subject of the construction, a construction that is presumably slightly more difficult, both because it is less usual and because, unlike (4), it includes two different arguments, a subject and a direct object. Example (6) also includes two arguments, through the addition of a *by*-agent. Note that passive LVCs with three arguments, namely a subject, the deverbal noun as direct object and a *by*-agent, are found in the native data (7) but not in the EFL data, probably because of the high complexity of this passive construction. Passive LVCs (of any of the above types) are found in native speech with a proportion of 6.1%. In EFL, the percentage rises from B1 (3.5%) to B2 (4.2%) and, especially, from B2 to C (6.6%). These results point to an increase in the syntactic complexity of the constructions as learners become more proficient. By the time they achieve a C level, they seem to have reached (and even slightly surpassed) the native standard in terms of passive use.

- (3) *you don't know a word you ma= you should er look it up a dictionary and you **make** a mistake you er mustn't be sad* (TLC-ITA-B1)
- (4) *I think the film was **made** in a small village in United States* (TLC-SPA-B1)
- (5) *he was afraid about erm being **given** a a different treatment or or or something like that* (TLC-SPA-C2)
- (6) *well erm the single most important I think erm was er the allegations **made** by the second runner-up in the election* (TLC-MEX-C2)
- (7) *what advice have you been **given** by your parents regarding your career?* (TLC-EXA)

5. Light verb constructions in EFL and ESL

In the second part of this study, the focus shifts to the possible influence of the context of acquisition (EFL or ESL) on the use and development of LVCs. The comparison relies on the EFL data analysed in Section 4 and ESL data representing Indian speakers of English, supplemented by the ENL data from the examiners. As was the case in the previous section, LVCs are considered as a category, but a distinction is also made between the three verbs (*MAKE*, *TAKE* and *GIVE*) for some of the analyses. Besides the frequency of LVCs, more formal aspects like syntactic complexity are briefly considered.

5.1. Frequency of light verb constructions in EFL and ESL

From a very general point of view, the ESL data are similar to the data previously examined in that they are not normally distributed, including many speakers who produced no LVCs (44% of TLC-IND), a majority of speakers who produced very few LVCs, and a few speakers who produced an unusually large number of LVCs (outliers). However, Figure 4, a boxplot representing the relative frequency of LVCs in the three varieties under study, viz. EFL, ESL and ENL, points to some intervarietal differences. This is confirmed by a Kruskal-Wallis test, which reveals a highly significant difference between the varieties ($H(2)=24.354$, $p<.001$). More precisely, a post hoc test shows that the difference lies between EFL and ESL, as well as between ENL and ESL. As was hypothesized in Section 2, LVCs, overall, are significantly more frequent in ESL ($Q3=1.69$, $Mdn=0.82$) than in EFL ($Q3=1.10$). It might thus be argued that ESL speakers' more regular exposure to English leads to a greater use of LVCs than is the case in an EFL environment. Interestingly, however, ESL speakers use even more LVCs than native speakers ($Q3=1.53$), resulting in a significant overuse of the construction in ESL. As already pointed out in Section 4.1, the difference between EFL and ENL is not statistically significant. These results might perhaps indicate that for ESL speakers the 'lexical teddy bear' effect claimed to pertain to high-frequency verbs (see Section 1) has spread to light verb uses as well, while for EFL speakers the lexical teddy bear effect might be limited to some of the more basic, less idiomatic uses, as suggested by Altenberg & Granger (2001).

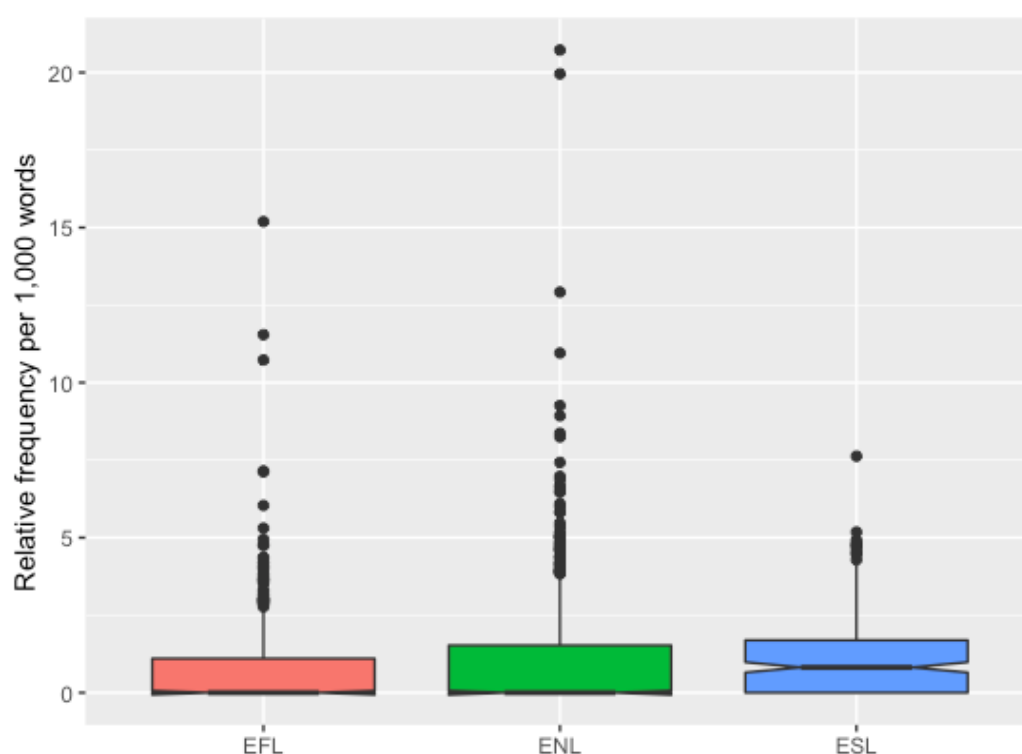


Figure 4. Relative frequency per 1,000 words of light verb constructions (all verbs) in EFL, ENL and ESL

Figure 5 shows that not all light verbs follow the same trend, though.¹³ Each of the Kruskal-Wallis tests carried out per verb points to a highly significant difference between the varieties ($H(2)=31.644$ for *GIVE*, 47.747 for *MAKE* and 31.619 for *TAKE*, $p<.001$), but not all pairs of varieties differ in a significant way, as revealed by a post hoc test, and those that differ

¹³ These statistics are based exclusively on speakers who produced at least one occurrence of LVC.

significantly do not necessarily display the same tendency (overuse or underuse). Only LVCs with *GIVE* are used significantly more in ESL ($Q3=1.19$) than in EFL ($Q3=0$). LVCs with *MAKE* display the opposite tendency, being used significantly less in ESL ($Q3=0.71$) than in EFL ($Q3=1.11$, $Mdn=0.60$). Those with *TAKE* do not differ significantly between ESL ($Q3=1.12$, $Mdn=0.63$) and EFL ($Q3=1.14$). In comparison with native speech, LVCs with *GIVE* are underused in EFL, as mentioned in Section 4.1 ($Q3=0$, to be compared with 1.46 in ENL), but they are neither underused nor overused in ESL ($Q3=1.19$). LVCs with *MAKE* were shown to be underused by EFL speakers ($Q3=1.11$, to be compared with 1.73 ($Mdn=1.00$) in ENL); they are also underused by ESL speakers ($Q3=0.71$). As for LVCs with *TAKE*, which were shown to be overused by EFL speakers ($Q3=1.14$, to be compared with 0.91 in ENL), they are also overused by ESL speakers ($Q3=1.12$, $Mdn=0.63$). It thus appears that each light verb can display different tendencies, as already demonstrated for EFL in Section 4.1. It also emerges from the results that, while the verbs tend to behave differently, quantitatively speaking, in EFL and ESL (with the exception of *TAKE*), they show similar tendencies in comparison with native speech (underuse of *MAKE* and overuse of *TAKE*; *GIVE* is an exception in this respect).

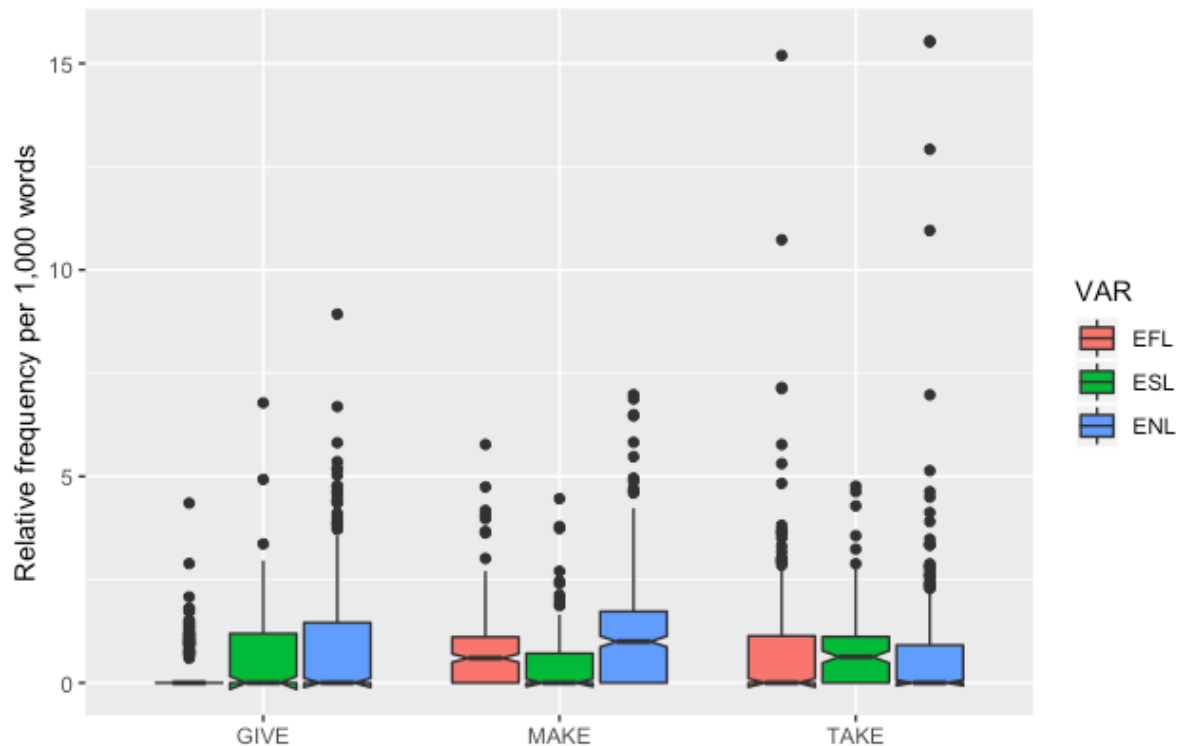


Figure 5. Relative frequency per 1,000 words of light verb constructions with *GIVE*, *MAKE* and *TAKE* in EFL, ESL and ENL

5.2. Frequency of light verb constructions across proficiency levels in EFL and ESL

If we combine the EFL/ESL comparison with the developmental perspective, we obtain Figure 6, a boxplot showing the relative frequency of LVCs in EFL and ESL across proficiency levels, with ENL as a reference. As already pointed out in Section 4.2, LVCs in EFL display a modest but steady rise as proficiency increases, from a $Q3$ of 0.99 at B1, through 1.09 at B2, to 1.46 ($Mdn=0.74$) at C, where they get close to (but never quite reach) the native standard ($Q3=1.53$). The evolution is less linear in the case of LVCs in ESL: $Q3$ amounts to 1.68 at B1, drops to

1.43 ($Mdn=0.87$) at B2, and then rises to 2.63 ($Mdn=1.13$) at C,¹⁴ which is higher than the native standard of 1.53. While the Kruskal-Wallis test for ESL is highly significant ($H(3)=33.774, p<.001$), it is only the difference between B1 and C ($B1<C$), and that between C and ENL ($C>ENL$) that are statistically significant, as revealed by a post hoc test. What is particularly characteristic of LVCs in ESL, therefore, is the increase in frequency between B1 and C, from a situation in which there is no significant difference between ESL and ENL, to a situation in which the ESL speakers overuse LVCs in comparison to ENL. The context of acquisition, and in particular the degree of exposure to English, thus seem to have an impact on the starting point and end point of the developmental path of LVCs in the EFL and ESL data from the TLC subsample: EFL speakers start and end below the native frequency, whereas ESL speakers start at the same level as native speakers, but end up surpassing them. In both contexts, however, the use of LVCs by proficient L2 speakers is problematic in that it differs from the native standard, being either underused (EFL) or overused (ESL).

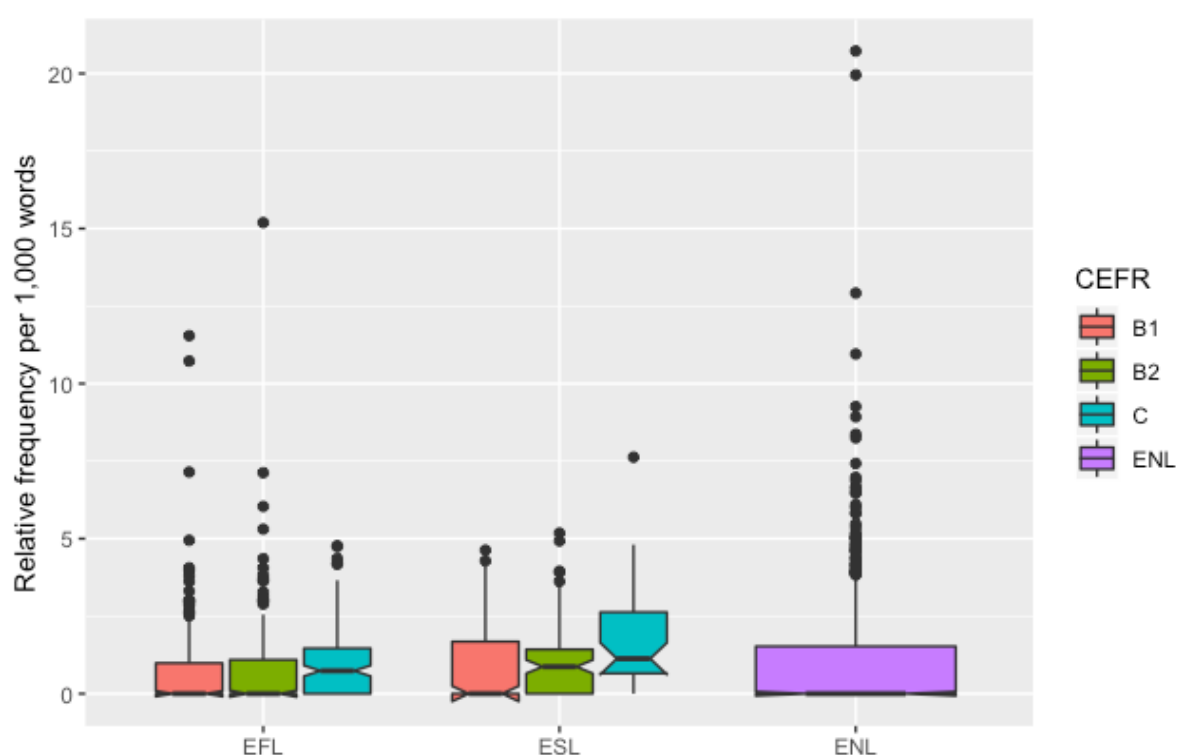


Figure 6. Relative frequency per 1,000 words of light verb constructions (all verbs) across EFL and ESL proficiency levels (B1, B2 and C) and in ENL

Finally, it must be noted that, as was the case in EFL, differences may emerge when a distinction is made between the different light verbs in ESL. The frequency of *GIVE*, for example, does not differ in any significant way across the CEFR levels nor between the CEFR levels and ENL. *MAKE* is only underused at B1 and B2 levels (the difference at C level is not significant). As for *TAKE*, it is overused at B2 and C levels. The differences between one CEFR level and the next are not significant for individual verbs, which might be due to the scarcity of data points at this fine-grained level of analysis, in a data set that is smaller than the EFL data set, being limited to a single linguistic and cultural background (Indian background).

¹⁴ The frequency of LVCs at the C level of ESL is the only case where the *IQR* and *Q3* differ from each other, as appears from the fact that, in Figure 6, this box starts above 0. The *IQR* here amounts to 1.97.

5.3. Formal aspects of light verb constructions in EFL and ESL

As shown previously for EFL only, interesting findings can also emerge from considering more formal aspects of LVCs. This will be illustrated in this section with the syntactic complexity of the constructions in EFL and ESL. Passivization, already discussed in Section 4.3 for EFL, is a feature that distinguishes the two varieties. With an average proportion of 11.5%, the passive voice in ESL is much more common than in EFL (4.9%) and, in fact, than in ENL (6.1%).¹⁵ The passive LVCs also tend to be more sophisticated in ESL than in EFL, with examples of three-argument passive constructions not found in the EFL data (Section 4.3), but occurring as early as the B1 level in the ESL data:

- (8) *I am er working on it and er I've been **given** many suggestions by my parents* (TLC-IND-B1)

Through premodification of the deverbal noun, LVCs can become more complex syntactically (and also more sophisticated lexically), as is clear from a comparison of (9) and (10). In EFL, premodified nouns represent a proportion of 10.8% of all LVCs. In ESL, the proportion rises to 19.1%. The greater complexity of LVCs in ESL also appears from the presence of coordinated deverbal nouns, as illustrated in (11). Although this is admittedly a very rare phenomenon, it is still slightly more common in ESL (2.3%) than in EFL (0.5%). Note that, additionally, the LVCs in ESL are more likely to combine different types of complexity. Thus, (12) is both an instance of a passive construction and a construction including two coordinated deverbal nouns, while (13) has two coordinated deverbal nouns, the second of which is premodified.

- (9) *it's impossible to **make** progress er with with them* (TLC-SPA-B2)
(10) *so I think the Indian government should at least **take** the few basic steps of providing them with some electricity and basic needs of some spaces* (TLC-IND-B2)
(11) *she has to make sure that she **gives** them all the love and caring which erm she claims to mm give every animal* (TLC-IND-C2)
(12) *he should be admitted somewhere **given** the g-gui= er guidance and counselling* (TLC-IND-C2)
(13) *we have a function called reception in which families and friends of the couple comes and **give** them blessing and go= er good wishes* (TLC-IND-B1)

6. Discussion

Using a newly available resource, the Trinity Lancaster Corpus, this study has explored the use of light verb constructions in L2 speech across different sections of the corpus, focusing on the proficiency level and the context of acquisition (EFL/ESL). The analysis of the EFL data has revealed a tendency among learners to underuse LVCs with *GIVE* and *MAKE*, and to overuse LVCs with *TAKE*, which indicates a potentially divergent use of LVCs depending on the verb they include, as already suggested for writing by previous corpus studies (Section 2). It has also shown that an increase in proficiency tends to go hand in hand with an increase in the

¹⁵ The higher frequency of the passive voice in ESL as compared to EFL seems to be a more general phenomenon, not limited to LVCs. Thus, Granger's (2013) findings for the passive in the Asian L2 English data from ICNALE show that the ESL populations use systematically more passive verbs than the EFL populations (though not more than the native reference corpora used in the study).

frequency of LVCs and a convergence towards the native reference. However, whatever the proficiency level, there is still a difference between the native standard and the EFL learners, which is statistically significant in the case of learners at the B1 and C levels. A more qualitative approach suggests that LVCs may evolve across proficiency levels in other respects than frequency, for example in semantic abstraction or syntactic complexity. Overall, these results seem to support the hypothesis that the use of LVCs among EFL learners tends to improve as proficiency increases, although even advanced learners diverge from native speakers in their use of the construction.

The structure of the TLC offers an unprecedented opportunity to compare EFL and ESL speech. In comparison with the EFL speakers, the ESL speakers appear to use more LVCs overall, presumably as a result of their more regular exposure to naturalistic English. While at first sight this might appear to support the hypothesis that the ESL speakers have a better knowledge of LVCs than the EFL speakers, it should be noted that they actually overuse LVCs in comparison to the native speakers (although the specifics vary depending on the verb). This suggests that ESL speakers, through the substantial input they receive in English, are more familiar with LVCs than EFL speakers and are thus affected by the ‘lexical teddy bear’ effect, or, more precisely, by what could be referred to as a ‘constructional teddy bear’ effect. An important caveat, in this respect, is that the ESL data set used here was limited to Indian speakers of English and that LVCs have been shown to be more frequent in Indian English than in any other South Asian English varieties (Hoffmann et al. 2011). With different / more varied ESL data, the frequency of LVCs might have been lower, and perhaps closer to the native values. If we examine developmental patterns, we notice that the overuse of LVCs is characteristic of more advanced ESL candidates; B1-B2 candidates use LVCs in more native-like frequencies. In terms of syntactic complexity, finally, the ESL candidates seem to have progressed further than the EFL candidates, and here too they sometimes surpass the native speakers themselves (cf. proportion of passive constructions).

As regards the spoken register of the data, we saw in Section 4.1 that the frequency of LVCs in EFL speech as compared to ENL speech did not seem to differ markedly from that in writing. Yet, this does not mean that LVCs in EFL are not affected by medium. While the many definitions found and applied in the literature make it difficult to compare findings from different studies (see Section 2), a comparison with results that were obtained for writing within the frame of another project but using the same criteria to identify LVCs (Gilquin 2007, 2010) shows that the constructions are more frequent in (French) EFL writing (13.82/10,000 words) than in EFL speech (7.30/10,000 words). It should be noted, however, that a similar difference appears to exist in native English (18.38/10,000 words in writing vs 8.61/10,000 words in speech), which might be an indication that, contrary to what is expected given the multi-word nature of LVCs, and contrary to what is usually claimed about these constructions (e.g. Quirk et al. 1985: 751), LVCs are perhaps not necessarily more typical of spoken than of written registers (see Smith 2009 or Shahrokny-Prehn & Höche 2011 for some evidence in support of this view). A comparison of our results with those obtained in Laufer & Waldman (2011), a developmental study of LVCs in (Hebrew) EFL and ENL writing, suggests that the developmental path of LVCs in EFL speech and writing might be quite similar: the authors of the study found an increase in frequency at the advanced level, still resulting in an underuse, but observed a “slow and uneven” development (Laufer & Waldman 2011: 664). Here again, however, the use of identical criteria to identify LVCs might bring out more fundamental differences between speech and writing. As for the impact of medium on EFL vs ESL, it might be expected to be stronger on the latter, with a better reflection of native stylistic differences, since users of ESL seem to be more sensitive to register differences than EFL learners, at least when it comes to multi-word units (see Gilquin 2015 on phrasal verbs). What would be needed,

however, in order to make claims about the real impact of register, is a large-scale study of LVCs in comparable EFL and ESL spoken and written data.

Finally, it is important to emphasize that the above results are not necessarily generalizable beyond the L2 speakers investigated here. For one thing, the selected TLC subsample may not be representative of EFL and ESL speech as a whole because of the specificities of the data (only one ESL population, bias towards EFL learners of Romance L1 background, etc.). For another, the corpus represents a special type of genre, namely examiner-candidate interaction, which might be different from normal spoken interactions and have an impact on the language that is produced, both by the examiner and the candidate. Within the frame of the present study, it thus appeared that examiners seemed to try to avoid periphrastic constructions with high-frequency verbs, preferring instead a simpler and more direct expression of the verbal meaning through a synthetic verb. Consider the following examples, in which the examiner starts with a high-frequency verb, presumably to be used in a periphrastic construction (light verb construction, causative construction, etc.), but then goes on to use a synthetic verb (in bold):

- (14) *so give me just **summarise** for me er briefly how the development of communication has helped erm human beings* (TLC-EXA)
- (15) *would you have made would you have **insisted** on any particular rules if your wife was a smoker and then you had children?* (TLC-EXA)
- (16) *if you get married to somebody erm and your wife and your mother have an argument would you take your would you **support** your mother or your wife?* (TLC-EXA)
- (17) *yes I take I c= I completely **understand** your point and agree with it too* (TLC-EXA)

The frequency of LVCs in native speech might thus be slightly underestimated. Taking the candidate's perspective, it turned out that a number of LVCs were topic-dependent, being elicited by a prompt from the examiner. This is exemplified in (18), where the examiner asks the candidate what advice she would give to an unhealthy friend. As a response, the candidate uses the LVC *give advice* twice. The frequency of LVCs produced spontaneously by the candidates might thus be slightly overestimated.

- (18) Examiner: *so if you if you had a friend who wasn't healthy wasn't fit*
Candidate: *mm*
Examiner: *what what would you say to her? what advice would you give to her?*
Candidate: *I will always give her advice to erm eat how much ever she wants but equally sleep <unclear/> these are the two advices I would like to give*

7. Conclusion

The analysis of LVCs in the TLC subsample has made it possible to provide some preliminary results on the (quantitative and qualitative) use and development of these constructions in spoken EFL and ESL. Both the candidate's proficiency level and his/her exposure to English seem to have an impact on the frequency and type of LVCs produced, although a more advanced level and/or more exposure to English do not necessarily lead to a native-like use. Advanced EFL and ESL speakers, for example, both differ in a significant way from native speakers, but while the former underuse LVCs, the latter overuse them.

Given the specificity of the data used in this study (see Section 6), the above findings would have to be confirmed against comparable EFL-ESL spoken data that represent speakers from a wider range of backgrounds and, ideally, expressing themselves in more spontaneous contexts. The findings should also be checked against similar findings obtained with a longitudinal design, following individual speakers over a long period of time. Other aspects of LVCs should be examined, in particular their standard or non-standard use (errors were found in both EFL and ESL, even at advanced levels of proficiency). The rich metadata included in the TLC could also be put to good account, especially to investigate individual profiles and to further specify proficiency levels (e.g. by taking exam achievement and task mark into account). Finally, the results should be revisited using more sophisticated, fine-grained quantitative methods. In particular, the univariate statistical comparisons drawn here should give way to multivariate analyses, which could consider the interaction between the different variables. It is to be hoped that, thanks to such follow-up studies, we will gain a better understanding of the use of LVCs by EFL and ESL speakers and that, ultimately, we will be able to help these L2 speakers get even closer to the native target, for, as pointed out by Sinclair (1990: 147), LVCs “contribute to the impression of fluency in English given by a foreign user”, and fluency is a skill that any student should want to improve.

Acknowledgements

I wish to thank the reviewers and the editors of this special issue for useful comments on an earlier version of this paper. Special thanks are also due to Samantha Laporte for her help with the statistics.

References

- Allerton, D. J. 2002. *Stretched Verb Constructions in English*. London and New York: Routledge.
- Altenberg, B. & Granger, S. 2001. “The grammatical and lexical patterning of MAKE in native and non-native student writing”, *Applied Linguistics* 22(2), 173–194.
- Biber, D., Gray, B. & Poonpon, K. 2011. “Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?”, *TESOL Quarterly* 45(1), 5–35.
- Brugman, C. 2001. “Light verbs and polysemy”, *Language Sciences* 23(4-5), 551–578.
- Bulté, B. & Housen, A. 2012. “Defining and operationalising L2 complexity”. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins, 21–46.
- Bulté, B. & Housen, A. 2015. “Evaluating short-term changes in L2 complexity development”, *Círculo de Lingüística Aplicada a la Comunicación* 63, 42–76.
- Butt, M. & Geuder, W. 2001. “On the (semi)lexical status of light verbs”. In N. Corver & H. van Riemsdijk (Eds.), *Semi-lexical Categories: The Function of Content Words and the Content of Function Words*. Berlin: Mouton de Gruyter, 323–370.
- Council of Europe. 2001. *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Crossley, S.A., Salsbury, T., McNamara, D.S. & Jarvis, S. 2011. “What is lexical proficiency? Some answers from computational models of speech data”, *TESOL Quarterly* 45(1), 182–193.
- Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

- Field, A., Miles, J. & Field, Z. 2012. *Discovering Statistics Using R*. London: Sage.
- Gilquin, G. 2007. "To err is not all. What corpus and elicitation can reveal about the use of collocations by learners", *Zeitschrift für Anglistik und Amerikanistik* 55(3), 273–291.
- Gilquin, G. 2010. "Language production: A window to the mind?". In H. Götzsche (Ed.), *Memory, Mind and Language*. Newcastle: Cambridge Scholars Publishing, 89–102.
- Gilquin, G. 2015. "At the interface of contact linguistics and second language acquisition research: New Englishes and Learner Englishes compared", *English World-Wide* 36(1), 91–124.
- Gilquin, G. 2016. "Discourse markers in L2 English: From classroom to naturalistic input". In O. Timofeeva, A.-C. Gardner, A. Honkapohja & S. Chevalier (Eds.), *New Approaches to English Linguistics: Building Bridges*. Amsterdam: John Benjamins, 213–249.
- Granger, S. 2008. "Learner corpora". In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook, Volume 1*. Berlin: de Gruyter, 259–275.
- Granger, S. 2013. "The passive in learner English. Corpus insights and implications for pedagogical grammar". In S. Ishikawa (Ed.), *Learner Corpus Studies in Asia and the World, Vol. 1: Papers from LCSAW2013*. Kobe: School of Languages and Communication, Kobe University, 5–15.
- Hasselgren, A. 1994. "Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with vocabulary", *Applied Linguistics* 4(2), 237–260.
- Hoffmann, S., Hundt, M. & Mukherjee, J. 2011. "Indian English - An emerging epicentre? A pilot study on light verbs in web-derived corpora of South Asian Englishes", *Anglia* 129(3-4), 258–280.
- Juknevičienė, R. 2008. "Collocations with high-frequency verbs in learner English: Lithuanian learners vs native speakers", *Kalbotyra* 59(3), 119–127. Available at <http://www.journals.vu.lt/kalbotyra/article/view/7599>.
- Kilgariff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1), 7–36.
- Kittigossin, R. & Phoocharoensil, S. 2015. Investigation into learning strategies and delexical verb use by Thai EFL learners. *3L: Language, Linguistics, Literature – The Southeast Asian Journal of English Language Studies* 21(2), 63–72.
- Langer, S. 2004. "A linguistic test battery for support verb constructions", *Linguisticae Investigationes* 27(2), 171–184.
- Larsen-Freeman, D. 1997. "Chaos/complexity science and second language acquisition", *Applied Linguistics* 18(2), 141–165.
- Laufer, B. & Waldman, T. 2011. "Verb-noun collocations in second language writing: A corpus analysis of learners' English", *Language Learning* 61(2), 647–672.
- Meunier, F. 2015. "Developmental patterns in learner corpora". In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 379–400.
- Mukherjee, J. & Hundt, M. (Eds.). 2011. *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a Paradigm Gap*. Amsterdam: John Benjamins.
- Neff-van Aertselaer, J. & Bunce, C. 2008. "Acquisition of *have*-constructions in the academic English of Spanish EFL university students: A comparative corpus study". LAUD Paper No. 718, Universität Duisburg-Essen. Available at <http://www.linse.uni-due.de/laud-paper.html>.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Nesselhauf, N. 2009. "Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties", *English World-Wide* 30(1), 1–25.

- Osborne, J. 2015. "Transfer and learner corpus research". In S. Granger, G. Gilquin & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 333–356.
- Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Scheepers, R. 2017. "South African students' use of delexical multiword units: The trouble with high-frequency verbs", *Stellenbosch Papers in Linguistics* 47, 89–114.
- Shahrokny-Prehn, A. & Höche, S. 2011. "Rising through the registers – A corpus-based account of the stylistic constraints on Light Verb Constructions", *Corpus* 10, 239–257. Available at <http://journals.openedition.org/corpus/2110>.
- Shirato, J. & Stapleton, P. 2007. "Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan", *Language Teaching Research* 11(4), 393–412.
- Sinclair, J. (Ed.). 1990. *Collins Cobuild English Grammar*. London: Collins Cobuild.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Smith, A. 2009. "Light verbs in Australian, New Zealand and British English". In P. Peters, P. Collins & A. Smith (Eds.), *Comparative Studies in Australian and New Zealand English: Grammar and Beyond*, Amsterdam: John Benjamins, 139–155.
- Szudarski, P. 2012. "Effects of meaning- and form-focused instruction on the acquisition of verb-noun collocations in L2 English", *Journal of Second Language Teaching and Research* 1(2), 3–37.
- Szudarski, P. & Carter, R. 2016. "The role of input flood and input enhancement in EFL learners' acquisition of collocations", *International Journal of Applied Linguistics* 26(2), 245–265.
- Traugott, E. C. 1999. "A historical overview of complex predicate types". In L. J. Brinton & M. Akimoto (Eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*. Amsterdam: John Benjamins, 239–260.
- Wang, D. 2011. "Language transfer and the acquisition of English light verb + noun collocations by Chinese learners", *Chinese Journal of Applied Linguistics* 34(2), 107–125.
- Wang, Y. 2016. *The Idiom Principle and L1 Influence: A Contrastive Learner-corpus Study of Delexical Verb + Noun Collocations*. Amsterdam and Philadelphia: John Benjamins.