COLLOCATIONS WITH HIGH-FREQUENCY VERBS IN LEARNER ENGLISH: LITHUANIAN LEARNERS VS NATIVE SPEAKERS

Rita Juknevičienė

Department of English Philology
Vilnius University
Universiteto g. 5, LT-01513 Vilnius
Tel.: +370 5 2687228
E-mail: rita.jukneviciene@takas.lt

Aim

Analysis of learner language, or interlanguage, produced by learners of different mother tongues has many implications for the theory and practice of teaching/learning English as a foreign language (EFL). The aim of the present research is to analyse collocational competence, of Lithuanian learners of English. The study was designed as a pilot study to analyse the learners' ability to produce collocations with high-frequency verbs, i.e. HAVE, DO, MAKE, TAKE and GIVE, and compare it with data from a comparable corpus of native speakers of English. It is assumed here that similar studies can shed light on the specific difficulties that Lithuanian learners of English face and contribute to the enhancement of teaching and learning outcomes.

Previous research

Collocational competence is often recognized as an important component of vocabulary acquisition (Nattinger and DeCarrico 1992, Lewis 1993, 2000, Woolard 2000), which might contribute to a better understanding of specific difficulties faced by learners of different mother tongues. In general, research into collocational competence is conducted in two directions. Some researchers focus on direct tests of collocations, e.g. Martynska (2004) analysed Polish EFL learners whereas, more recently, Jaén (2007) reported findings from a test of collocations administered to native speakers of Spanish. The other research direction is concerned with the investigation of data extracted from corpora of authentic learner language. One of the largest research projects in the area, i.e., compilation of a large corpus of learner language, representing different mother tongue backgrounds, was initiated by S. Granger. The resultant International Corpus of Learner English (ICLE) was used in a number of contrastive studies. Collocations, in particular, were analysed by Altenberg and Granger (2001) who among other aspects described collocations with the

delexical MAKE produced by Swedish- and French-speaking learners. Their study showed that errors of collocations can often be attributed to the influence of the mother tongue (ibid., pp. 179-180). A study focused on the influence of German as L1 was undertaken by Nesselhauf (2005), who provided an elaborate description of collocational errors and reported that about a quarter of the collocations produced by German learners are wrong whereas one third are deviant or questionable (2005, 237) while the errors often arise from non-congruence of collocations in German and English.

Research into the interlanguage of Lithuanian learners is rather scarce. One of the early publications belong to R. Aprijaskytė and E. Pareigytė (1982) who described frequent lexical errors of Lithuanian learners of English. More recently, Žindžiuvienė (2003) presented a paradigm of teaching essay writing in English. Her survey of Lithuanian learners showed that over 80 per cent of the learners indicated their deficient lexical competence as a major obstacle to successful writing. It is therefore expected that the Lithuanian component of the ICLE corpus (LICLE), which is being currently compiled in Vilnius University (Bikelienė, Grigaliūnienė and Juknevičienė *forth.*), will encourage more research into the specificity of teaching/learning EFL in the Lithuanian context.

The definition of collocations used in this paper draws on the phraseological rather than statistical or frequency-based approach. Proposed by A. P. Cowie and developed by Howarth (1998), this classification of word phrases applies several criteria (Howarth 1998, 27), of which lexical substitution or commutability is often considered to be the most relevant (Nesselhauf 2005, 27). The commutability of a word phrase is related to possible lexical substitution of its elements, which gives base for the following grouping of word phrases: free combinations, e.g. *want a car, read the paper*, collocations, e.g. *shrug one's shoulders, make a decision*, and idioms, e.g. *sweeten the pill, kick the bucket* (Nesselhauf 2005, 32-33). Collocations, differently from free combinations and idioms, are defined as word combinations having arbitrary restriction on the commutability of their elements. Thus, the verb in *make a decision* cannot be replaced by the synonymous *do* or *produce*. It is in this sense that the term 'collocation' will be used throughout this article.

The current analysis of collocations is seen as a first step in the description of collocational competence of Lithuanian learners of English and is limited to a certain type of collocations. Firstly, it is concerned with *verb + noun (object)* collocations with such verbs which have the highest frequency in the English language, i.e. HAVE, DO, MAKE, TAKE and GIVE. Secondly, the analysis focuses on the delexical use of the verbs when they combine with direct objects and build phrases whose meaning largely depends on the meaning of the noun. In grammars of English these combinations, e.g. *to make a decision*, *to give an example* etc., are termed 'delexical structures' (Sinclair 1990, 147-151) or 'light verb uses' (Huddleston and Pullum 2002). Semantically such combinations fall under the phraseological definition of collocations because their meaning is transparent; yet substitutability of the components is semantically restricted, e.g., *to take notice*, but not **to take observation*, *to make a statement*, but not **to do a statement*.

It is restrictions in the substitutability of components that seem to cause difficulty to EFL learners. As a consequence, EFL learners avoid collocations and instead 'rely on larger,

rarer, and clumsier words which make their language sound stilted and awkward' (Sinclair 1991, 79). Furthermore, combinations with delexicalised words are less likely to exhibit correspondence in translation and thus are more error-prone in learner language (Lewis 1993, Nesselhauf 2005). The analysis of delexical combinations with MAKE by Swedish and French learners of English showed that EFL learners tend to avoid such collocations in writing even though they significantly overuse high-frequency verbs in comparison to native speakers (Altenberg and Granger 2001: 174). Yet Biber et al. (1999, 1027-1029) found that the English high-frequency verbs when used delexically form quite many collocations typical of the academic register and newspaper language, both of which are often seen as end-goals of EFL or EAP (English for Academic Purposes) study programmes. It is therefore expected that this study will shed light on the specific difficulties Lithuanian learners face when acquiring academic English vocabulary and contribute to a better understanding of the development of collocational competence.

Data and methods

Collocations for the analysis were extracted from two sources, i.e. corpora of native speaker (NS) and non-native speaker (NNS) learner essays. The NS corpus is the LOCNESS corpus, compiled by S. Granger and often used as a reference corpus in contrastive studies of learner language. For the purpose of this analysis, only part of LOCNESS corpus was used so that the size and structure of the NS corpus would match the composition and topics of the NNS corpus[1]. The NNS material was extracted from the LICLE corpus. Table 1 presents the structure and size of the corpora.

Table 1. ***Learner corpora used in the study***

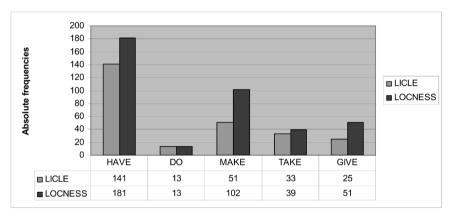|	**LICLE (NNS corpus)**	**LOCNESS (NS corpus)**
Learners	Lithuanian university students (3rd-4th year)	British and American university students
Number of essays	139	103
Total corpus size	78640 words	90412 words
Average essay length	566 words	875 words

The NNS corpus includes essays written by native speakers of Lithuanian who indicated in the questionnaires that Lithuanian is their mother tongue. About 20 per cent of the corpora texts were literary essays, accounting for 13107 words in LICLE and 19019 in LOCNESS. The other essays in both corpora were argumentative essays on a variety of topics related to different linguistic, social or moral issues[2].

---

[1] The following subcorpora of LOCNESS were included in the analysis: brsur2.cor , brsur3.cor, mrq0001.1-46.1

[2] A full list of topics of LICLE essays is published in Bikelienė, Grigaliūnienė and Juknevičienė (forth.). The LOCNESS essay topics are listed at http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/LOCNESS1.htm

As explained above, the current analysis is focused on the delexical uses of the verbs HAVE, DO, MAKE, TAKE and GIVE when they form collocations of the pattern *verb+noun (object)*, e.g., *to have impact, to do research, to make a decision, to take an advantage,* and *to give an example.*

Word combinations matching the above defined criteria were selected manually. The manual selection, however, was eased by the Wordsmith Tools software (v. 5) to compute concordances of individual verb lemmas. The concordance lines were scrutinized to eliminate irrelevant cases, e.g., HAVE and DO are often used as auxiliaries, MAKE can be used as a causative verb as in *to make a conference interesting*. The total number of selected collocations was 263 in the LICLE essays and 386 in the LOCNESS essays. Graph 1 gives numbers of collocations in the two corpora. The analysis included only such collocations which were represented by at least two occurrences.

	HAVE	DO	MAKE	TAKE	GIVE
LICLE	141	13	51	33	25
LOCNESS	181	13	102	39	51

Graph 1. **Distribution of collocations with delexical verbs in the two corpora**

A special attention in the selection process was given to possible lifting of collocations from the task prompt used for the examination essays as some of the prompts are rather lengthy quotations of ca. 100 words. It turned out that the Lithuanian learners did indeed employ the vocabulary of the prompts in their own writing. The following collocations were found in the wording of the prompts and student essays: *to make a move* (occurs eight times in student essays), *to make an analysis* (five occurrences), *to take interest* (two occurrences). All the instances of these collocations were examined to establish if they could have been lifted from the prompts. It was found that the collocations occurred only in the essays written to the prompts which included the collocations themselves and they do not occur once in essays written on 'collocation-free' prompts. Another point to confirm lifting is that the collocations were usually incorporated in the passages which paraphrased the prompt, i.e. the opening/concluding paragraphs to state the author's view on the topic at hand. It was thus decided to exclude these collocations from further analysis.

122

Results

On average, LICLE essays contain fewer than two collocations per essay whereas the LOCNESS essays have 3.74. Hence NS students produce almost twice as many collocations with delexical verbs as the NNS learners. Even though collocations are underused in the NNS writing, a closer analysis revealed that differences in the absolute frequencies are significant only for some verbs. Table 2 presents Log Likelihood values (statistically significant values are printed in bold) which describe differences in the frequency of collocations in the two corpora. In the table, the minus sign '-' indicates underuse of the collocations in LICLE in relation to LOCNESS, whereas '+' indicates overuse (Rayson 2004).

Table 2. **Log Likelihood values (p < 0.05, critical value 3.84)**

	Collocations with HAVE	Collocations with DO	Collocations with MAKE	Collocations with TAKE	Collocations with GIVE
LICLE vs. LOCNESS	(-0.97)	(+0.13)	**-10.96**	(-0.01)	**-5.82**

Statistically significant differences were established in the frequency of collocations with the verbs MAKE and GIVE, both of which are underused in the NNS language. Owing to the relatively small size of the corpora under analysis, statistical differences have not been expected to be very big. Moreover, statistical analysis reveals only a general tendency and it does not take into account acceptability of certain collocations. Therefore, qualitative analysis was undertaken to highlight other differences. A table of ten most frequent collocates of each of the five verbs is given in Appendix 1.

The qualitative analysis revealed that Lithuanian learners often misuse collocations. Decisions on the acceptability or correctness of collocations produced by the EFL learners in similar studies are usually based on native speaker intuitions, dictionaries and English corpus material. In this study, correctness of collocations was supported by evidence in the British National Corpus (BNC) and two dictionaries, i.e. the *Oxford Dictionary of Collocations* (2002) and the *Collins Cobuild English Dictionary for Advanced Learners* (2001). Interestingly enough, Nesselhauf reports that the judgement of native speakers, whom she used in her study, yields contradicting results as they often disagree about the acceptability of certain collocations (2005, 43-53). Since this study is rather limited in scope, native speaker judgement was only resorted to in exceptional cases when no other support could be obtained.

The analysis of wrong or deviant collocations in LICLE showed that transfer from the mother tongue seems to be a prevailing strategy and English collocations are often created by translating them word for word from Lithuanian, e.g.,

(1) *Statistics show that over 3 years we lost about 15% of Lithuanians, which lets us <u>make sad conclusions</u>: Lithuanians are escaping their motherland as a land of Black Death. (corrected: ... which allows to come to a sad conclusion / to conclude that ...)*

The noun *conclusion* in often associated with cohesion of the academic essay and is evidently learnt as a connective device, e.g. *In conclusion*, which perhaps explains why the learners feel safe when using this well-known word in combinations with other words. Its collocability in English, however, is different from Lithuanian and should be therefore specifically addressed in teaching. In the present study, **to make a conclusion* is one of the most frequent collocations with MAKE in the LICLE essays yet it never occurs in ICLE and is very rare in the BNC—only 15 instances were found (cf. 4698 of *to make a decision*).

Translation from Lithuanian lead to another L1 specific error which is related to the learners' inability to differentiate between MAKE and DO, which are both rendered in Lithuanian by one word, i.e. DARYTI. Hence, the corpus includes the following unacceptable collocation:

(2) *Therefore they found language a construct, "played" certain language games and made linguistic experiments*. (corrected: …did experiments…)

Sometimes the strategy of translating from L1 leads to creation of collocations which are typical of argumentative texts in Lithuanian but are very rare in ICLE, for example,

(3) *to have a possibility / problems / experience / differences.*

In terms of statistical difference, collocations given in (3) are significantly underused in LOCNESS or do not occur even once. They are, however, acceptable English collocations well-represented in the BNC. Possibly, the underuse is due to the limited scope of material of the study or specificity of vocabulary of the analysed corpora. Evidently, Lithuanian learners construct their argumentation by relying on lexis transferable from their L1 which may be different from NS argumentative vocabulary and which leads to a different quality of language.

Another source of deviant or misused collocations is related to the inadequate mastery of English, or so-called developmental factors. Quite a number of questionable collocations have been created by the Lithuanian learners by combining high-frequency verbs, which are known to be used as safe 'lexical teddy bears' (Hasselgren in Altenberg and Granger 2001, 174), with a seemingly useful noun included in the essay prompt. In some cases this is done to refer back to the topic and develop argumentation, sometimes just to create a word combination. The following sentences illustrate the case:

(4) *The reform should be made not only to our University but also to our mentality.* (corrected: We should reform not only our …)

(5) *The fact is that Estonia had the right to do the transfer*. (corrected: …to transfer [the monument].)

The nouns *reform* and *transfer* were used in the prompt and then lifted by the students to create original collocations which are rare in the BNC and which never occur in LOCNESS.

Another distinctive group of collocations, which is specific to the Lithuanian corpus and significantly underused in the native speaker essays, is related to simple collocations which

are learnt very early at school which clearly belong to the core lexis of EFL learners. This is particularly obvious in the case of TAKE collocates, e.g., *to take place / care / photos*:

(6) *There is no doubt that certain reform should* <u>take place</u> *for the Lithuanian educational system not to collapse.* (better: …be undertaken, implemented, introduced, put in place etc.)

Admittedly, these collocations also occur in the ICLE essays yet statistically they are significantly overused by NNSs. Perhaps writing under stress and willingness to produce correct language is one of the reasons behind this phenomenon. It also points to the importance of learning/teaching collocations as semi-fixed word combinations. The students seem to feel safe when using *to take place* yet they are not aware of the fact that academic English has many phrases of similar fixedness. In comparison, they do not produce collocations with abstract nouns which are more typical of the academic register and which are significantly overused in LOCNESS, e.g., *to take a stand / decision / effect*.

The inadequacy of academic vocabulary of Lithuanian learners is particularly obvious in the use of collocations with MAKE. Even though the number of different words (types) is almost similar, the NNS learners significantly underuse collocations with MAKE, which in fact confirms findings of other similar studies. Altenberg and Granger (2001) report that a substantial proportion of collocates extracted from the native speaker corpus are nouns denoting speech acts or otherwise related to speaking, i.e. *claim*, *argument, point, case, statement, assumption,* and *reference*. Though the current analysis used only certain components of the LOCNESS corpus, so-called speech collocates in this study account for 38% of all collocates of MAKE used by NSs. In comparison, Lithuanian learners significantly underuse such nouns (*statement*, 3 occurences, or 5% of all collocates). Furthermore, they significantly fall behind both Swedish and French learners who created many more collocations with speech nouns, i.e. 13% and 9% of all MAKE collocations respectively (Altenberg and Granger 2001: 180). Collocations with speech nouns are more formal stylistically and thus more difficult to master to foreign learners of English. Hence, explicit teaching of academic English and its typical collocations should be considered.

Conclusions

The analysis of collocations with the high-frequency verbs HAVE, DO, MAKE, TAKE and GIVE confirms findings of similar contrastive studies investigating NNS and NS collocational competence and shows that there are both quantitative and qualitative differences in the use collocations with high frequency verbs between Lithuanian learners of English and native speakers. The results of the study show that, on the one hand, Lithuanian learners significantly underuse collocations typical of the academic register, which may be explained by their deficient academic vocabulary. To compensate it, they resort to translation from L1 when creating collocations. This strategy, however, often leads to misuse or creation of word combinations which are rare in English. On the other hand, Lithuanian learners seem to be aware of collocations and their contribution to the quality and fluency of the text and thus try to use collocations that they know very well or even create new

word combinations to make their texts better. It is therefore possible to conclude that to achieve better results the teaching of academic English might be specifically focused on typical collocations of the register.

As a pilot study, the analysis was limited in scope so its findings point to several further research directions. Firstly, no attempt has been made here to investigate grammatical accuracy of the collocations, e.g. the use of determiners, prepositions, transformations of the *verb+noun* collocations etc. Arguably, research into the so-called *local phrase grammar* might reveal more qualitative differences between NNS and NS learners. Secondly, if collocational competence does develop by stages, it might be worthwhile comparing the LICLE material with the language produced by lower-level learners to get a better understanding of factors behind the successful acquisition of collocations.

REFERENCES

Aprijaskytė R., Pareigytė E., 1982. *Some lexical difficulties for the Lithuanian learner of English*. Vilnius: VU leidykla.

Altenberg B., Granger. S., 2001. The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing. *Applied Linguistics 22/2:* 173-195.

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

Bikelienė L., Grigaliūnienė J., Juknevičienė R. *forthcoming*. The Lithuanian Component of the International Corpus of Learner English (LICLE): a resource for English language learning, teaching and research at Lithuanian institutions of higher learning.

Howarth P., 1998. Phraseology and second language proficiency. *Applied Linguistics,* 19 (1), 24-44.

Huddleston R., Pullum G. K. (ed.)., 2002. *The Cambridge Grammar of the English Language*. Cambridge: CUP.

Jaén M. M., 2007. A corpus-driven Design of a Test for Assessing the ESL Collocational Competence of University Students. *International Journal of English Studies* 7(2). Murcia: Servicio de Publicaciones. 127-147.

Lewis M., 1993. *The Lexical Approach*. London: Language Teaching Publications.

Lewis M. (ed.), 2000. *Teaching Collocation. Further developments in the lexical approach.* Hove: Language Teaching Publications.

Martynska M., 2004. Do English language learners know collocations? *Investigationes Linguisticae,* vol. XI, 2-12.

Nattinger J. R., DeCarrico J. S., 1992. *Lexical Phrases and Language Teaching.* Oxford: OUP.

Nesselhauf N., 2005. *Collocations in a Learner Corpus*. Amsterdam: J. Benjamins Publishing Company.

Rayson P., 2004. Log-likelihood calculator. Available at www.ucrel.lancs.ac.uk.

Sinclair J. (ed.), 1990. *Collins Cobuild English Grammar*. Glasgow: Harper Collins Publishers.

Sinclair J., 1991. *Corpus, Concordance, Collocation.* Oxford: OUP.

Woolard G., 2000. Collocation – encouraging learner independence. *Teaching Collocation. Further developments in the lexical approach* ed. by M. Lewis, 28-46. Hove: Language Teaching Publications.

Žindžiuvienė I., 2003. *Synthetic Paradigm of Teaching/Learning Compositions in a Foreign Language*. A summary of PhD thesis in English. Kaunas: VDU.

DIDELIO DAŽNUMO ANGLŲ KALBOS VEIKSMAŽODŽIŲ KOLOKACIJOS BESIMOKANČIŲJŲ KALBOJE: LYGINAMASIS LIETUVIŲ IR GIMTAKALBIŲ KALBOS TYRIMAS

Rita Juknevičienė

S a n t r a u k a

Straipsnyje pristatomas anglų kaip užsienio kalbos besimokančiųjų lietuvių studentų kolokacinės kompetencijos tyrimas. Tyrimo medžiagą sudaro *veiksmažodžio-daiktavardžio (tiesioginio papildinio)* sintaksinio modelio kolokacijos, išrinktos iš argumentacinių rašinių tekstyno (LICLE), kurį sudaro pažengusiųjų lygio rašiniai. Šio tekstyno tyrimo rezultatai lyginami su medžiaga, surinkta išanalizavus panašaus pobūdžio gimtakalbių anglų kalbos vartotojų rašinių tekstyną (LOCNESS). Medžiagos analizė rodo, kad gimtakalbių ir negimtakalbių kolokacijų vartojimas skiriasi tiek kiekybiškai, tiek kokybiškai. Gimtakalbių tekstyne rastos kolokacijos dažniau vartojamos akademinėje anglų kalboje, o lietuvių studentų rašiniuose dažniau ir taisyklingiau vartojamos sakytinei anglų kalbai būdingos kolokacijos. Kolokacijos vartojimo klaidos, rastos LICLE tekstyne, dažniausiai randasi dėl gimtosios kalbos įtakos.

APPENDIX 1. *Ten most frequent collocates of each high-frequency verb in learner language (in brackets – the number of occurrences)*

	LICLE (NNS)	LOCNESS (NS)
HAVE	right (20), ideas (8), power (7), time (7), impact (6), possibility (6), problems (6), ability (5), choice (5), influence (5)	right (23), effect (14), children (13), power (11), sex (11), consequences (7), strength (6), time (6), chance (5), impact (5)
Types	34	41
Tokens	141	181
DO	harm (3), job (4), task (2), transfer (2), work (2)	job (5), research (4), study (4)
Types	5	3
Tokens	13	13
MAKE	sense (7), conclusion (6), decision (6), change (5), step (5), impact (3), reform (3), statement (3), use (3), career (2)	decision (39), claim (9), argument (8), point (7), mistake (6), case (5), comparison (5), statement (5), assumption (3), attempt (3)
Types	16	15
Tokens	51	102
TAKE	place (9), action (4), part (4), position (4), care (3), look (3), advantage (2), attitude (2), photos (2)	advantage (8), place (7), stand (4), action (3), decision (3), care (2), effect (2), part (2), position (2), precedence (2)
Types	9	12
Tokens	33	39
GIVE	opportunity (4), possibility (4), arguments (3), birth (3), pleasure (3), chance (2), education (2), information (2), speech (2)	chance (8), example (4), information (4), opportunity (4), reason (3), rise (3), time (3), account (2), effect (2), hope (2)
Types	9	17
Tokens	25	57