

DOI: <https://dx.doi.org/10.18454/RULB.11.12>

Камшилова О.Н.

Доцент, кандидат филологических наук,

Российский государственный педагогический университет им. А.И. Герцена

ПЕРЕПРОИЗВОДСТВО В ИНТЕРЬЯЗЫКЕ: ЧАСТОТНЫЙ И ГРАММАТИЧЕСКИЙ АСПЕКТ*Аннотация*

Исследование основывается на данных корпуса текстов петербургских школьников, изучающих английский язык (SPbEFL Learner Corpus). В статье рассматривается характерное для интеръязыка явление перепроизводства в количественном (частота) и качественном (грамматическая корректность) аспектах. Лингвистический анализ двух базовых грамматических структур – *S V Od* и *S V C* со связкой *be* – фокусируется на выборе дополнения в первом случае и дополнения к подлежащему во втором. В результате доказываем, что перепроизводство данных структур характеризуется не только их частотностью в речи обучающихся, но и специфическим выбором их составляющих.

Ключевые слова: корпус ученических текстов, интеръязык, перепроизводство, базовые грамматические структуры.

Kamshilova O.N.

Associate professor, PhD in Philology, Herzen State Pedagogical University of Russia

OVERUSE IN LEARNER LANGUAGE: FREQUENCY AND ACCURACY*Abstract*

The investigation bases on the SPbEFL Learner Corpus findings and addresses the issue of overuse in learner output in relation to both frequency and accuracy of basic grammar structures. The article deals with two basic syntactic structures – *S V Od* and *S V C* with “*be*” as copula. The linguistic analysis concerns the choice of object in the first case and subject complement in the second. The objective is to show that overuse of the structures in question results not only in their frequency in learner production, but also in learner specific choice of their constituents.

Keywords: learner corpus, interlanguage, overuse, basic grammar structures.

Почта авторов / Author Email: onkamshilova@gmail.com

Introduction

Since the end of the 1990-s learner corpora (LC) – electronic collections of written and spoken texts, produced by L2-learners – have been regarded as a most relevant resource for learner language studies. LC studies have pinpointed most typical mistakes of learners with different L1, which caused productive changes in learning materials and teaching techniques [4], [5], [6], [10]. Corpus-based studies proved also helpful for interlanguage analysis [2], [10].

Saint Petersburg EFL Learner Corpus (SPbEFL LC), based upon the LC design criteria [5, P.8], is a comparatively small multi-L1 (Russian, Chinese, Japanese, Korean, Thai, and Vietnamese) corpus.

The contributors to the corpus were 90 high school students (average age 15.4) from Saint Petersburg (Russia) and their 12 peers, new immigrants to CA, USA (average age 15.5) with the pre-tested intermediate (26%) and upper-intermediate (74%) language proficiency.

The corpus contains written texts (essays and personal letters), monologues and dialogues in scripts. The genres and the topics were suggested by the school syllabus and the format of the State General Exam (ЕГЭ).

Task setting was deliberately different from the requirement of S. Granger’s corpus (ICLE¹): the text production was timed and a size limit for written text was set. No reference tools (dictionaries or grammars) were used either.

Thus, the study presumes, that the contributors’ texts are as near as spontaneous, and the learner output will positively demonstrate interlanguage strategies, such as overuse or underuse.

Method

Any corpus study is a method by definition, since it is based on application of corpus managers, tools that produce such relevant information as concordance, word counts, frequencies, collocation, and syntactic patterns.

Comparison of different corpora is used to pinpoint

specific data for SPbEFL LC authors.

The investigation bases on the assumption that both the vocabulary and the sentence patterns presumably reflect the actual language fund – interlanguage or interim language – that the learners subconsciously resort to in case of FL communication (*cf.* [3], [9]). The method applied was comparing L2 with L1 varieties [2], [4], [8], [10].

Discussion

Learner language in SLA research is described as developing a transfer grammar (interlanguage) with *overuse*, *underuse*, and *fossilization* as learner strategies. Overuse and underuse as learner language characteristics are obviously a matter of frequency. An investigation of basic grammar structures in L2 learner speech and a closer exploration of their fill provided evidence for at least three interpretations of overuse:

- overuse 1 – learners use a word / construction *A* *more frequently* than native speakers (NSs);
- overuse 2 – learners use a word/construction *A* *instead of* a word / construction *B*;
- overuse 3 – learners use a word / construction *A* *with a learner specific cast*.

Following the hypothesis, corpus data can pinpoint the most frequent grammatical structures, used by L2 learners. The question is how much they resemble the Basic English Grammar structures.

The list of Basic Grammar constructions taken as model in this paper is adopted from [1]:

- *S V A* (Mary is in the house);
- *S V C* (Mary is kind / a nurse);
- *S V Od* (Somebody caught the ball);
- *S V Od A* (I put the plate on the table);
- *S V Oi Od* (She gives me presents);
- *S V* (The child laughed).

We shall focus on two of them, namely *S V C* with *be* as copula and *S V Od* with *have* as main verb.

Results

A comparative frequency list analysis of three raw learner corpora – French [6, P.17], Quebec [2] and the SPb EFL LC – is only a rough exploratory survey, but it provides some interesting perspectives.

Table 1 – Top 10 word forms in three LC

	French LC	Quebec LC	SPbEFL LC
1	the	the	I
2	of	to	to
3	to	I	and
4	a	a	you
5	and	of	the
6	is	and	a
7	in	in	is
8	that	that	it
9	it	is	in
10	be	it	have

The learner language proficiency varies in the corpora from advanced in the French and Quebec LCR to intermediate / upper-intermediate in SPbEFL LC. Besides, the task setting criteria were different: the contributors to SPbEFL corpus were set a time limit and did not use any reference materials so that the text production was *nearly* spontaneous (except for the previous class practice). Therefore, the rich vocabulary and developed sentence patterns trained in class would give way to simple, common lexis and transparent structures.

The top 10 in the compared corpora suggest that the task subjects in SPbEFL corpus were definitely 1st-person oriented, hence the first rank of *I*. The attraction here is the high frequency rank of *and*, as well as the verb forms *have* and *is*.

As the concordance displays showed, the conjunction *and* is used to connect short and numerous clauses, homogeneous parts, to start a sentence and to fill the pause in case of hesitation – all that features in spontaneous speech production.

In the comparable corpora the only verb forms in the top 10 list are the forms of *be* which can be either lexical, link, modal or auxiliary. The high frequency of *have*, which is mostly used as a lexical verb, suggests that SPb schoolchildren make wide use of the pattern I/WE <HAVE> N.

This pattern analysis found 489 hits for *have* in the whole SPbEFL LC. Lexical use of *have* was found in 429 contexts. Modal use is high enough – 52, while the use of auxiliary *have* (for Perfect forms) is insignificant (8).

The low rank of auxiliary *have* can be definitely marked as underuse: the learners avoid perfective and progressive forms. This avoidance is a failure to nuance the aspect of the event / situation described.

What seems quite special for the learner texts is the use of the basic grammar construction S V Od where the Od position is filled with nominalized forms. That is, the learners prefer a precast pattern to a more conventional for NSs adjectival complement (*be free, be independent*) or verbal predicate (*to communicate / talk*):

1. *If you leave your childhood house you'll have your own life²*
2. *If you have your own accomodation you also have a freedom*

3. *I enjoy to have communication with interesting people from different countryes*

As examples (1-3) demonstrate, learners resort to S <HAVE> Od construction *instead of* constructions S V C and S V, which we previously defined as “overuse 2”. This type of overuse is a roundabout way of expressing the learner’s idea, a help they find from what they are familiar with. Thus, frequency of S <HAVE> Od in learner output is not overproduction alone, but also a strategy to find a way out.

Still more evidence for treating overuse as a multivalued feature was found while searching the SPbEFL corpus for S V C constructions with *be*-copula. It is common knowledge that this type of construction is overused by EFL speakers [7]. The data in table 2 prove the general tendency: the majority of *be*-forms found in the essays are S <BE> C cases, while only a few contexts realize their auxiliary and lexical use. The underuse of auxiliary *be* is another proof of avoiding analytical verb forms (progressive and passive).

Table 2 – The use of BE in SPbEFL (Essay)

BE functions → BE forms found↓	Copula verb	Auxiliary verb (progressive forms)	Auxiliary verb (passive)	Lexical Verb (There is/are)
is	230	3	1	1
's	67	1	-	-
was	10	1	2	-
am	1	-	-	-

NS corpus data argue that S <BE> C construction is very popular in native speech, too. It is important, that the typical complements here differ in important ways across registers:

“Over 50% of the complements of *be* copula are noun phrases. This structure is extremely common, occurring about 10,000 times per million words (or several times on every page of prose)” [1, P. 446].

“The copula *be* is overwhelmingly the most common verb taking an adjectival complement, occurring over 20 times more than any other copular verb. *Copular be + adjective* occurs over 5,000 times per million words for all registers (more than twice per page on average). This pattern is especially common in academic prose and fiction” [*ibid.*, P.437].

The above statements, backed by LWSEC³ findings, pinpoint that

- NSs use S <BE> C construction frequently;
- most frequent complements are noun phrases (NP);
- adjectival complements (AP) are preferable in academic prose and fiction;
- frequency of the construction in learner speech may not be the only value of overuse.

Comparison of SPbEFL LC data with those of LWSEC shows that complements of *be* copula in Russian learner output are preferably adjectival (59%), though the register in corpus is mainly, even in the Essay part, conversational (table 3):

Table 3 – Subject predicative realization in SPbEFL LC

	AP	NP	PP	Complement clause	Inf P
S <BE> C (1325)	779	475	18	40	3

The choice of the complement adjective intensifies the “academic touch” in Russian learner subcorpus, as they seem to favor *difficult* and *different*, both marked for academic register in NS corpus [*ibid.*, P.440]. So, the overuse of S <BE> C is definitely accompanied with a learner specific cast, in this particular case due to L1 interference and imperfect teaching practice and textbooks:

4. *work / period (about youth) is difficult*
5. *It is difficult to meet people who don't know English / to keep pets / to find a good job.*

Similarly, the choice of noun complement specifies the overuse of the construction in a functional aspect (table 4).

Table 4 – Comparing predicative nouns frequency

	SPbEFL	LWSEC (conversation)
1	proper name	crap
2	idea	proper name
3	thing	home
4	friend	no way
5	problem	people
6	webster	matter
7	film	thing
8	place	time
9	child/children	trouble
10	dog	way

The learner choice is for NPs that identify the logical class or type to which the subject belongs (descriptive use), while most common NP subject predicatives in NS corpus “are attitudinal, marking the stance of the speaker / writer” [*ibid.*, P. 450].

So, in both cases overuse is not a matter of frequency alone: while the wrong choice of adjectival complement suggests a learner specific cast (“overuse 3”) and demonstrates lack of accuracy, the noun complement, generally correct, is used in a limited set of functions.

Corpus findings also revealed a common learner error which may be treated as mal-use rather as it starts as a subject complement structure S <BE> C but essentially it is a basic structure with a verbal predicate S V Od or S V, and the copula can even precede modal verbs (6-8):

6. *I'm totally agree with them.*
7. *I'm prefer pizza, meat, fish and others*
8. *I'm study in art-school*

This error frequency is remarkable and it occurs across all types of texts, written and spoken. What is more, it is found in the output of learners with different L1⁴:

9. *... it's depend on how much people have a good responsibil (CA Essay)*
10. *But, in any case it's depend on person (SPb Essay)*
11. *... it's can be right some people who work well...(CA Essay)*
12. *I hope that it's wouldn't last for long time (SPb Monologue)*

13. *As for me I'm always go to internet. I'm play online games (SPb Essay)*

14. *Yes, sometimes I'm go to the cinema with my friends (SPb Dialogue)*

Conclusion

The research was intended to provide evidence for a multivalued character of overuse in learner language. It is suggested in the paper, that overuse is not a matter of frequency alone and that there are at least three interpretations of it.

A high frequency rank, higher than in NS production (“overuse 1”) may mark register specific features. Thus, the high rank of the conjunction *and* in SPbEFL corpus is explained as a regular spontaneous speech production feature: connecting short and numerous clauses, homogeneous parts, starting a sentence and filling the pause in case of hesitation.

Overuse may be referred to the use of a precast pattern instead of some other pattern or construction (“overuse 2”), thus finding help from familiar patterns, violating their accuracy, but expanding their nominative and functional properties. This is the case with S <HAVE> Od basic construction, which learners often resort to instead of constructions S V C and S V. So, S <HAVE> Od frequency in learner output is not overproduction alone, but also a strategy to find a way out.

Sometimes the frequency of a construction in learner output may be comparable to that in native speech and it is only a quantitative and qualitative analysis of its constituents that reveals an important difference in their choice, which marks inaccuracy of both composition and function. In Russian learner output subjective predicative constructions (S V C) with *be*-copula are preferably adjectival. Adjectival complements are common for academic register in native speech, while in conversation NP complements are preferable. The choice of the complement adjective itself intensifies the “academic touch” in Russian learner output.

The repertoire of noun complement displays a difference that specifies the overuse of one function: the learner choice of NPs is intended for descriptive use, while most common NP subject predicatives in NS corpus mark the stance of the speaker (attitudinal use).

So, S <BE> C overuse is not a matter of frequency alone, either: the wrong choice of adjectival complement suggests a learner specific cast (“overuse 3”) and demonstrates lack of accuracy; the noun complement, generally correct, is used in a limited set of functions.

Overuse is often accompanied with inaccuracy. This can be proved by a found common learner error which, likely resulting from the overuse of S <BE> C construction, may be treated as mal-use rather as it starts as a subject complement structure being essentially a basic structure with a verbal predicate S V Od or S V.

¹International Corpus of Learner English: <https://uclouvain.be/en/research-institutes/ilc/cecl/icl.html>

²Here and after the illustrations of learner text from SPbEFL LC are given with their authentic spelling and grammar preserved

³LWSEC – Longman Written and Spoken English Corpus

⁴SPb marks learners from S. Petersburg with Russian as L1, CA is for learners from California with different East-Asian L1s

Список литературы / References

1. Biber D. Longman Grammar of Spoken and Written English / D. Biber, S. Johansson, G. Leech and others. – Harlow: Longman, 1999. – 1204 p.
2. Cobb T. Analyzing late interlanguage with learner corpora: Quebec replications of three European studies / T. Cobb // Canadian Modern Language Review. – 2003. – Vol.59 (3). – P. 393-423.
3. Corder S.P. Error Analysis and Interlanguage / S.P. Corder. – Oxford: Oxford University press, 1981. – 120 p.
4. Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitsprachenerwerbsforschung / M. Walter, P. Grommes (eds.) – Tübingen: Max Niemeyer Verlag, 2008. – 211 p.
5. Granger S. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching / S. Granger. – Amsterdam: J. Benjamins Pub. Comp., 2002. – 246 p.
6. Granger S. The computer learner corpus: A versatile new source of data for SLA research / S. Granger // S. Granger (ed.) Learner English on Computer. – London: Longman, 1998. – P. 3-18.
7. Hinkel E. Simplicity without elegance: Features of sentences in L1 and L2 academic texts / E. Hinkel // TESOL Quarterly. – 2003. – Vol. 37(2). – P. 275-301.
8. Mukherjee J., Schilk M. Verb-complementational profiles across varieties of English: Comparing verb classes in Indian English and British English / J. Mukherjee, M. Schlik // T. Nevalainen, I. Taavitsainen, P. Pahta, M. Korhonen (eds.) The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present. – Amsterdam: John Benjamins, 2008. – P.163-181.
9. Selinker L. Interlanguage / L. Selinker // International Review of Applied Linguistics. – 1972. – Vol. 10. – P. 209-231.
10. Tono Y. Learner corpus research: Some recent trends / Y. Tono // G. Weir, S. Ishikawa (eds.) Corpus, ICT, and Language Education. – Glasgow: University of Strathclyde Publishing, 2010. – P. 7-17.

Список литературы на английском / References in English

1. Biber D. Longman Grammar of Spoken and Written English / D. Biber, S. Johansson, G. Leech and others. – Harlow: Longman, 1999. – 1204 p.
2. Cobb T. Analyzing late interlanguage with learner corpora: Quebec replications of three European studies / T. Cobb // Canadian Modern Language Review. – 2003. – Vol.59 (3). – P. 393-423.
3. Corder S.P. Error Analysis and Interlanguage / S.P. Corder. – Oxford: Oxford University press, 1981. – 120 p.
4. Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitsprachenerwerbsforschung [Advanced learner varieties: corpus linguistics and second language acquisition studies] / M. Walter, P. Grommes (eds.) – Tübingen: Max Niemeyer Verlag, 2008. – 211 p. [in German]
5. Granger S. Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching / S. Granger. – Amsterdam: J. Benjamins Pub. Comp., 2002. – 246 p.
6. Granger S. The computer learner corpus: A versatile new source of data for SLA research / S. Granger // S. Granger (ed.) Learner English on Computer. – London: Longman, 1998. – P. 3-18.
7. Hinkel E. Simplicity without elegance: Features of sentences in L1 and L2 academic texts / E. Hinkel // TESOL Quarterly. – 2003. – Vol. 37(2). – P. 275-301.
8. Mukherjee J., Schilk M. Verb-complementational profiles across varieties of English: Comparing verb classes in Indian English and British English / J. Mukherjee, M. Schlik // T. Nevalainen, I. Taavitsainen, P. Pahta, M. Korhonen (eds.) The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present. – Amsterdam: John Benjamins, 2008. – P.163-181.
9. Selinker L. Interlanguage / L. Selinker // International Review of Applied Linguistics. – 1972. – Vol. 10. – P. 209-231.
10. Tono Y. Learner corpus research: Some recent trends / Y. Tono // G. Weir, S. Ishikawa (eds.) Corpus, ICT, and Language Education. – Glasgow: University of Strathclyde Publishing, 2010. – P. 7-17.