



2017

## The Development of Syntactic Complexity in the Writing of Russian Language Learners: a Longitudinal Corpus Study

Olesya V. Kisselev

Anna A. Alsufieva

Follow this and additional works at: <https://scholarsarchive.byu.edu/rlj>



Part of the [Slavic Languages and Societies Commons](#)

### Recommended Citation

Kisselev, Olesya V. and Alsufieva, Anna A. (2017) "The Development of Syntactic Complexity in the Writing of Russian Language Learners: a Longitudinal Corpus Study," *Russian Language Journal*: Vol. 67: Iss. 1, Article 3.

Available at: <https://scholarsarchive.byu.edu/rlj/vol67/iss1/3>

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Russian Language Journal by an authorized editor of BYU ScholarsArchive. For more information, please contact [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

## The Development of Syntactic Complexity in the Writing of Russian Language Learners: a Longitudinal Corpus Study

OLESYA V. KISSELEV

ANNA A. ALSUFIEVA

### Introduction

To make inferences about how second language (L2) learners develop over time, most Second Language Acquisition (SLA) research has traditionally relied on cross-sectional one-time sampling design, in which data collected from different groups of learners at different levels of language proficiency are compared against a preestablished set of measures. Rarer are longitudinal studies, in which researchers track a small number of participants over a relatively long period of time. Recent developments in technology and the rise of the language corpora have made it possible to combine the benefits of these two approaches; longitudinal *LEARNER LANGUAGE CORPORA*, large databases collected continuously from a group of learners over an extended period of time (i.e., semester, year, or program), successfully combine “longitudinal designs with dense developmental data collection” (Vyatkina 2012), and thus hold promise of providing a more fine-grained picture of the ebbs and flows of language development.

The study presented in this paper is an exploratory investigation of writing development conducted on the materials of a longitudinal corpus of learners of Russian. It attempts to explore the development of syntactic complexity in the writing of Advanced learners of Russian, as they moved from the Intermediate to the Advanced level of language proficiency as defined by the ACTFL proficiency scale.<sup>1</sup> The study focuses on the development of general complexity (as measured by sentence

---

<sup>1</sup> The ACTFL proficiency scale, created by the American Council on the Teaching of Foreign Languages (ACTFL), includes four main language proficiency levels: Novice, Intermediate, Advanced, and Superior. The ACTFL scale provides a thorough description of the linguistic skills and functions that are required to fulfill tasks at these different levels and is widely used in placement and testing of foreign language learners (see <https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012>).

length, sentences per 1,000 words in text, number of sentences per paragraph, and other sentence-based indices) and specific complexity, namely subordination and coordination (measured through the number and types of sentences with coordination and subordination).

The article is organized as follows. We first briefly describe the advances made in the field of learner corpus research and the benefits of corpora studies to language pedagogy. We then provide a short description of the Russian Learner Corpus of Academic Writing used in the current research. The next section presents a discussion of the construct of syntactic complexity. It is followed by a description of the measures of syntactic complexity chosen for this study, as well as the description of procedures and the subsequent analysis. Next, we report on the results and discuss the observations regarding the development of the dimensions of writing complexity in our data. The last section presents implications for pedagogical practice, as well as implications for future research.

### **1. Learner corpora studies**

Learner corpus research, an area of applied language studies that investigates large and systematically compiled collections of texts (oral or written) produced by second language learners, has grown exponentially in the past twenty years. This tremendous interest in learner corpus studies is credited to a large extent to the groundbreaking project created by Sylvaine Granger, The International Corpus of Learner English (ICLE, Granger 2003). ICLE, compiled of a large collection of essays written by advanced learners of English as a Foreign Language (EFL) of various language backgrounds, was used in a number of important large-scale studies (see, for example, Dagneaux et al. 1998; Granger 1996; Granger 1999; Leńko-Szymańska 2008; Nesselhauf 2005, 2005; O'Donnell et al. 2013; Gries and Wulff 2013; inter alia), which set out to establish universal and language-specific patterns of EFL acquisition. This line of research has effectively proposed a new model of analyzing learner language, the one that promises to help deliver a comprehensive linguistic description of linguistic abilities of language learners from different language backgrounds and at different developmental levels and, overall, to better our understanding of the processes and mechanics of language learning.

Learner corpora strive to supplement linguistic data with rich sociolinguistic information about the learners who contribute the data and the settings in which the data are collected; this information may include such parameters as age of acquisition and age of performance, linguistic experience and current linguistic level, instructional settings, and task characteristics. Learner corpora naturally hold particular interest for language instruction, since they can help associate learner characteristics and pedagogical events with emergence of particular linguistic structures (Belz and Vyatkina 2008; Leaver and Shekhtman 2002).

Despite their clear benefits, learner corpora representing languages other than English are rare, which limits the number and diversity of corpus-based studies in SLA and the pedagogy of these languages. This is especially true for less commonly taught languages such as Russian. To the best of our knowledge, at the time of writing this article, the only available systematically collected and complete Russian learner corpus is the Russian Learner Corpus of Academic Writing (RULEC).<sup>2</sup>

RULEC is a longitudinal corpus that consists of written texts produced by foreign language (FL) and heritage language (HL) learners of Russian who were enrolled in the Russian Flagship Program, a special program for advanced study of Russian at Portland State University. The advance track of the program admits students who are at least Intermediate-Mid level, and some students (many of them of HL background) start the program at Advanced-Low. The program offers a series of sequenced content-based courses in the Russian language, in which vocabulary, grammar, and syntax are mostly treated indirectly. More formal instruction targets such linguistic skills as forming written and oral paragraphs, essay writing, and oral presentation, as well as conducting research in Russian and writing a research paper. Writing assignments that the students typically have to respond to target a particular text function or communicative purpose, such as providing an argument for a position. Following the ACTFL guidelines, the text

---

<sup>2</sup> RULEC is available in the web-based format as a part of the Russian Learner Corpus (RLC, <http://www.web-corpora.net/RLC/>). RLC, which is currently under active development, is intended to include a number of different sub-corpora (such as RULEC), which represent various first language backgrounds and proficiency levels.

functions targeted in instruction are definition, paraphrase, summary, narration, description, expository writing, comparison and contrast, cause and effect, supported opinion, argumentation, and hypothesis. Blended types of these functions are also represented in the writing of RULEC authors, especially in research papers. In addition to a clearly defined textual function or goal of communication, the writing task also provides the intended audience (real or imaginary), such as readers of a weekly local newspaper or high school students in a partner high school Russian program. An example of an assignment may be to write a short chapter on water preservation for the sixth graders in the Russian-English immersion program, to provide a definition of a term central to the student's major, or to write an opinion letter to the newspaper editor as a reaction to a feature article.

Given the description of the types of assignments in RULEC, we suggest that these texts fit both broad and narrow definitions of academic genre. The broad definition, such as Hyland's (2007), recognizes any genre common in the academic community as academic writing. In the tradition of genre pedagogy, academic genre is defined through an orientation to the creation of the text. If academic literacy is defined as a set of skills that allow the student to analyze, critically evaluate, and utilize information for the purposes of the specific occasion, then an academic text is a text that demonstrates these skills (Coe 2002; Korotkina 2011). All texts in RULEC are common college writing assignments created as a reaction to presented information, with a clearly defined goal, purpose, and audience.

In the compilation of the corpus, all written assignments produced by the students were regularly collected over a period of four years. Since the majority of the students completed the program in two to three years, the corpus provides a very thick longitudinal data for all the corpus authors.

One of the unique features of RULEC is its detailed metadata: each text entered into the corpus includes information about the learner (such as gender, language background, and language level) and the text (such as the course for which the paper was written, targeted linguistic function, and time limit). The presence of metadata allows for an easy automatic grouping of all texts based on the various parameters; for instance, one can quickly create sub-corpora of FL and HL learner texts, or sub-corpora

based on the learners' ACTFL levels or text genres. Comparing sub-corpora can help researchers answer questions regarding the influence of early exposure, language level, or even language task on certain linguistic features. Due to the nature of texts available in RULEC and the corpus design, the corpus lends itself nicely to exploration of writing at advanced levels. The study described below is an attempt to analyze the development of students' writing as learners move from the Intermediate to Advanced level.

The indices of advanced writing are numerous, but generally researchers studying L2 writing agree that the quality of a written text can generally and aptly be assessed on the measures of lexical and syntactic complexity, accuracy, and fluency. Briefly, lexical complexity refers to expert use of vocabulary as reflected in vocabulary richness, lexical density, and lexical sophistication (Lu 2012). The construct of syntactic complexity is described below.

## **2. Syntactic Complexity**

Syntactic complexity is defined in research literature as a range of basic and sophisticated structures available and accessible to the learners, as evidenced in their oral or written production (Wolfe-Quintero et al. 1998; Ortega 2003). Syntactic complexity has been well established as an important construct and developmental index in L2 writing and speaking, and as such, as a valid assessment measure (Wolfe-Quintero et al. 1998; Ortega 2003; Osborne 2011; Lu 2011; Yang et al. 2015; Lu and Ai 2015). Examples of production units that have been traditionally used to measure syntactic complexity are clauses, T-units, and sentences; in addition to simple frequencies, the ratio of these units (e.g., the number of clauses per sentence, the number of dependent clauses per T-unit vs. coordinated clauses per T-unit, etc.), as well as their length, diversity, and accuracy, is also considered to be a metric of syntactic complexity. Lu (2011), in fact, identifies more than thirty syntactic complexity measures proposed in previous research, although not all of them receive equal attention in various studies.

Earlier studies of indices of syntactic complexity in the studies of development of L2 written communication—summarized and reviewed in Wolfe-Quintero et al. (1998)—tended to focus on T-unit measures. Based on their review, Wolfe-Quintero et al. (1998) concluded that the

number of clauses per T-unit and the number of dependent clauses per total number of clauses or T-units serve as the best indicators of syntactic complexity, because these structures appear to “exhibit a linear relationship to proficiency level across studies that used a wide range of levels” (Wolfe-Quintero et al. 1998, 99). The over-reliance on T-unit-based categories in the assessment of writing has been reevaluated in the more current literature (Lu, 2011; Biber et al. 2011); these researchers argue that although the increase in subordination may be indicative of moving from lower levels of writing ability to Intermediate levels, Advanced writing is characterized by nominalization strategies. Lu (2011, 56), for instance, after investigating fourteen different indices of syntactic complexity in English, concludes that the best measures in higher-level writing are complex nominals per clause and complex nominals per T-unit, complex phrases per clause and complex phrases per T-unit, as well as mean length of clause, sentence, and T-unit.

Yet, when attempting to gauge the developmental trajectories of learners’ writing skills, tracking a variety of indices of syntactic complexity rather than a small set of indices may be more advantageous. Norris and Ortega (2009), for instance, strongly advise applying a multidimensional approach to measuring syntactic complexity, in which GENERAL complexity measures (such as T-units, clauses, sentences, as well as ratios of these units) are to be supplemented by a variety of SPECIFIC complexity measures, including coordination, subordination, and particular phrases (e.g., complex nominals, verb phrases). Given that the field of Russian language acquisition can only rely on a very small number of empirical studies that could suggest which specific indices of linguistic complexity may be indicative of which proficiency level, a multidimensional approach to the study of Russian learner language is especially relevant. We suggest that our field stands to benefit from studies in which different general dimensions of language complexity (such as length of a sentence or a T-unit, the number of T-units per sentence or clause) can be supplemented by the investigation of specific complexity measures (such as specific syntactic structures).

### **3. Syntactic complexity in Russian**

Literature in the field of Russian as a Foreign Language research, which concerns itself with the issues of Russian academic discourse, provides

little guidance regarding which specific linguistic measures may be indicative of the genre; on the one hand, it emphasizes complex—mostly subordinated—sentences as a means of relating complex abstract ideas in a logical and continuous discourse (Barykina et al. 1978; Khimik 2003; Lobanova and Slesareva 1980). On the other hand, the preponderance of nouns and noun phrases in the Russian academic genre has also been noted (Prokhorova 1998). Russian academic writing is also said to be saturated with participles, adverbs, and, to a lesser degree, subjectless predicates. We recognize that academic prose in the definition of these Russian authors tilts towards more strictly speaking scientific discourse, but in the absence of other points of comparison, we suggest that the observations of these Russian as a Foreign Language specialists may be a useful point of departure. The question of how well Russian language learners can produce these complex structures is left largely unexplored in the current research literature; the few available studies that grapple with the question of Advanced-level discourse in learner Russian focus on Russian as a heritage language (RHL). These studies observe that the lower-level RHL speakers rely more on subordination (as evident in the larger proportion of T-units) than the higher-level RHLs and the native speakers (Dengub 2012) and that the RHL learners are less likely to use verbal adverbs and participles than the native speakers of Russian (Friedman and Kagan 2008; Dengub 2012). Building on these observations, the current study sets out to investigate the dynamics of coordination and subordination patterns in the writing of RULEC heritage and non-heritage learners as they move from Intermediate to Advanced Russian language proficiency.

#### **4. The study**

##### **4.1. Data**

The data for this project come from the Russian Learner Language of Academic Writing (RULEC) described above. For the study, we extracted texts authored by a cohort of sixteen students who had started the program at Intermediate level and progressed to Advanced level. The level of proficiency of these students was established based on the results of the Russian Flagship Overseas Qualifying Tests (implemented by the American Councils of International Education), which assess reading, listening, writing, and oral proficiency. Once the results of the tests were

obtained, the newly assigned level was used in the metadata supplementing the papers collected from this student-author from that point on. The sub-corpus extracted for the study includes both HL learners of Russian (eight subjects) and FL students (eight subjects). We excluded texts that consisted of less than a paragraph, as well as final drafts of research papers since the final drafts were corrected and edited by the instructors and tutors. The resulting corpus was further subdivided into four small sub-corpora on the basis of language background and language level: FL Intermediate, HL Intermediate, FL Advanced, and HL Advanced.

## 4.2. Measures

In choosing the particular measures of syntactic complexity, we had to deal with constraints of working with a large but raw, i.e., non-annotated, corpus. RULEC is not syntactically parsed, making it impossible to automatically extract indices of syntactic complexity, such as T-units, dependent clauses, noun phrases, etc. Nonetheless, even raw corpora allow for (semi)automatic analyses of many syntactic patterns, when the task is approached creatively. Below we describe the specific linguistic features that we used as indices of syntactic complexity.

4.2.1. General complexity. Sentence length—as calculated by the number of words per sentence, as well as a number of sentences per 1,000 words—was chosen as a general complexity measure following the arguments presented in Vyatkina (2012). Vyatkina (2012) explains that while T-units have to be coded manually, a task that can prove prohibitive when dealing with large volumes of data, sentence length can be generated automatically. More importantly, a sentence, compared to a T-unit, is a more psychologically real construct, and as a unit “directly produced by [a] learner” may more accurately reflect the learner’s intention in constructing the text (Bardovi-Harlig, 1992, 391, as cited in Vyatkina 2012, 582). In our paper, we also consider a related measure of number of paragraphs and paragraph length in words as an additional measure indicative of general writing ability.

4.2.2. Specific complexity. The specific syntactic complexity measures chosen in this study are coordination and subordination. To automatically extract coordinated and subordinated complex sentences, we searched the corpus for specific conjunctions and conjunctive words

(or pronominal words). Please note that for the sake of simplicity, we refer to all of these conjunctive devices as conjunctions in the article. The limitation of extracting sentences for analysis based on the presence of a conjunction is the loss of conjunction-less complex sentences; however, it is likely that the percentage of those in our data is relatively small: Russian language learners may in fact prefer to mark the logical relationships between clauses overtly through the use of conjunctions (Dengub 2012). Additionally, the focus on specific conjunctions, as opposed to a more general category of clause, allows us to take a closer look at types of conjunctions used by learners, range of forms, and patterns of their usage. We also examine the appropriateness, degree of sophistication, and fluency with which students use conjunctions at different levels.

#### 4.3. Conjunction as a unit of syntactic complexity

Conjunctions represent one of the primary means of signaling logical relationships between units in a sentence; they illustrate various semantic associations among elements of discourse, explicate the logical relations between parts of discourse such as causation, addition, succession, and contrast, and contribute to the cohesion and coherence of text (Halliday 1985). Although text cohesion and logic can be achieved through lexical means only (thus resulting in connector-less sentences), Russian academic prose in particular is marked by explicit, overt stress on logic and the development of an argument (Lobanova and Slesareva 1980). Marking the development of arguments overtly may be the reason for a preponderance of complex, poly-predicative sentences, in which elements (clauses) are connected by various conjunctions (in addition to linking adverbials and embedded phrases). Traditionally, conjunctions are divided into two types: coordinators (also called coordinating conjunctions, e.g., *и* “and,” *но* “but”) and subordinators (also called subordinating conjunctions, e.g., *что* “that,” *потому что* “because,” *после того как* “after”). Although conjunctions can connect intraclausal elements, in this paper we only focus on those instances where conjunctions connect clauses within a sentence.

#### 4.4. Research questions

Based on the chosen measures, the research questions guiding the present study are framed as follows: Is there an overall change in the amount,

functional variety, and accuracy of complex sentences with conjunctions in the writing of Advanced learners of Russian as compared to their performance at the Intermediate level? Is there a difference in the trajectory of writing development in regard to complex sentences with conjunctions between the heritage and non-heritage learners of Russian?

#### 4.5. Data analysis

The four sub-corpora, Intermediate FL, Intermediate HL, Advanced FL, and Advanced HL, were first subjected to unit statistics test, using the corpus analyzer program WordSmith Tools 6.0 (Scott 2014). This analysis provided general statistical information regarding such units as the number of word tokens, word types,<sup>3</sup> sentences, and paragraphs, as well as the information on the mean length of words, sentences, and paragraphs. The overall descriptive statistics on the four sub-corpora are presented in Table 1.

The information in Table 1 that is most relevant for our discussion of GENERAL SYNTACTIC COMPLEXITY is the number of sentences per 1,000 words, sentence length in words, and paragraph length in words. Sentence length grows slightly for both FL and HL learners as they move up the level, but at the same time the number of sentences per 1,000 words is decreasing. This is, of course, the same trajectory observed from two vantage points; as students write longer sentences, the proportion of longer sentences per 1,000 words has to decrease. This general trend suggests that the informational density of sentences is growing, which might be a good marker of writing development in general. Paragraph length also shows a slight increase as the learners progress towards more advanced language proficiency.

Since the numbers presented in Table 1 can only provide a panoramic look at the four sub-corpora, other manipulation of the data was needed. First, using the WordSmith Tools program, we created four wordlists from the four sub-corpora. We then searched the wordlists for the presence of coordinating and subordinating conjunctions used by the

---

<sup>3</sup> Word token is every occurrence of every word in the corpus; if a text is one hundred words long, it is said to contain one hundred word tokens. A word type is a distinct word form (*thing* and *things* are two types). Normally, some words in a text are repeated because they are particularly frequent in the language or because they are topically salient; thus, a text that is one hundred words long may contain only sixty distinct types.

learners. After the extraction, we ran concordance searchers for each conjunction in the corpus. We manually separated conjunctions from conjunctive words and introductory phrases, and assessed the appropriateness of the use of each conjunction in the particular context and evaluated the grammatical accuracy of the structure in which the conjunction was involved.

*Table 1. Overall text statistics across four sub-corpora*

Units	FL Intermediate	FL Advanced	HL Intermediate	HL Advanced
Number of Texts	318	462	289	510
Words (tokens)	58,236	94,903	44,177	107,916
Types (distinct forms)	12,857	15,639	11,131	18,983
Mean length of text in words	181.42	205.42	152.85	211.6
Number of sentences	4,507	6,805	3,199	7,254
Number of sentences per 1,000 words	77	72	72	67
Sentence length in words	12.92	13.95	13.81	14.88
Paragraphs	516	623	339	627
Number of sentences per paragraph	8.7	10.9	9.4	11.5

In all, the corpora contained forty-seven conjunctions, all can be considered typical in the language produced by native speakers if compared to the relative frequency of different conjunctions in standard Russian based on the data from the word frequency dictionary of the Russian language by Lyashevskaya and Sharoff (2009). The relative order of frequency of the conjunctions in our data approximates that of the native speakers; for example, *čto* “that” is the most frequent conjunction in the above-mentioned dictionary, as well as in the writing of RULEC learners.

We then organized the forty-seven conjunctions into groups depending on their functional characteristics and normalized the frequencies of the conjunctions, prorating them per 1,000 sentences.<sup>4</sup>

#### 4.6. Coordinating conjunctions

Thirteen coordinating conjunctions of varying frequencies were found in the data; their distribution and frequencies are presented in Table 2.

*Table 2. Frequencies of coordinating conjunctions across groups and levels prorated per 1,000 sentences*

Coordinating conjunction	Intermediate FL	Intermediate HL	Advanced FL	Advanced HL
<i>и</i> "and"	130.4	24.2	97.1	30.2
<i>а</i> "and/but"	37.5	26.3	23.2	34.7
<i>либо</i> "or"	2	0.4	2.2	2
<i>но</i> "but"	74.1	36.5	58.8	45.9
<i>однако</i> "however"	19	3.2	7.9	3.4
<i>зато</i> "though"	0	0.4	0.5	0.6
<i>то есть</i> "that is"	12.3	7	7.1	9.2
<i>а именно</i> "namely"	0.8	0.4	0	0.1
<i>такой как</i> "same as"	19.6	29.1	11	33.6
<i>не столько ... сколько</i> "as much"	0	0	0	0.6
<i>по мере того как</i> "at the time"	0	1.4	0.3	2.2

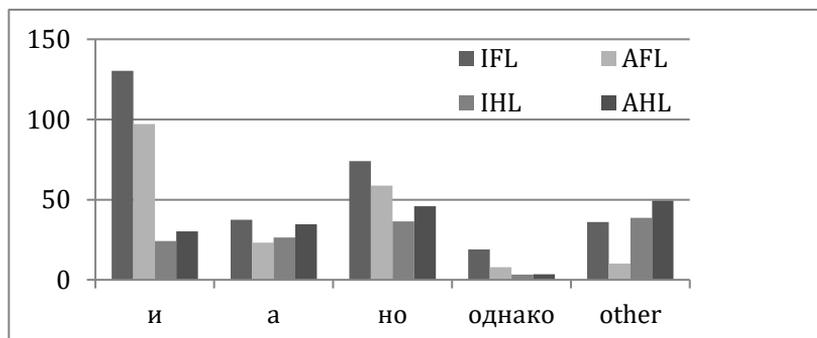
---

<sup>4</sup> Given the differences in size in the number of tokens and in the number of sentences between the four sub-corpora, the normalization of each conjunction per 1,000 sentences was carried out. Such normalized counts allow us to directly compare the possible differences in the number of conjunctions between the four sub-groups of learners.

<i>в то время как</i> “as soon”	0.8	0.4	0	0.1
<i>тогда как</i> “while”	0.6	0	0	0.6
Total	166.7	105.1	111	133

The coordinating conjunctions in our data represent the following functional categories: copulative conjunctions: *и* (cf. and); disjunctive conjunctions: *либо* (cf. if); adversative conjunctions: *а* (cf. but, and), *но* (cf. but), *однако* (cf. however), *зато* (cf. although); explanatory conjunctions: *то есть* (cf. that is), *а именно* (cf. in other words), *такой как* (cf. the same as); and contrastive-comparative conjunctions: *не столько . . . сколько* (cf. not so much as), *в то время как* (cf. at the same time), *по мере того как* (cf. along the way). Unsurprisingly, copulative, disjunctive, and adversative conjunctions are more numerous than the other types of coordinating conjunctions, which are grouped together in Figure 1 below. In terms of distributions of the various conjunctions of this type in the learner data, we see a decrease in the use of these conjunctions in the writing of FL learners and a slight increase in the writing of HL learners. This tendency is expected: as FL writers acquire more and more new structures, their reliance on the “basic” coordinating conjunctions lowers. In the case of HL writers, they appear to improve their ability to express their thoughts in writing in general and begin to more frequently overtly mark the relations between ideas expressed in different clauses.

Figure 1. Coordinating conjunctions by functional group



In addition to obtaining numeric information, we also coded all sentences with conjunctions for accuracy. In this analysis, we considered any deviation from the standard Russian as an error. Errors included missing or misplaced punctuation (example 1), the choice of wrong or ineffective conjunction (examples 2 and 3), or structural issues, such as lack of grammatical agreement (example 4).

(1) он не согласен с тем фактом \***что**, россияне в царской России жили хорошо.

he does not agree with the fact that Russians in the tsarist Russia lived well. (HL learner)

он не согласен с тем фактом, **что** россияне в царской России жили хорошо. (Standard Russian)

(2) \***Поскольку** я понимаю экономику, использования иностранной рабочей силы оказывает позитивное влияние на социально-экономическое развитие России.

**Since** I understand economics, the use of foreign work force has a positive influence on the development of socio-economic situation in Russia. (FL learner)

**Насколько** я понимаю экономику, использования иностранной рабочей силы оказывает позитивное влияние на социально-экономическое развитие России.

**As far as** I understand economics, the use of foreign work force has a positive influence on the development of socio-economic situation in Russia. (Standard Russian)

(3) . . . ставит пьесы не только в России, \***а** в Европе.

. . . directs shows not only in Russia, **but** in Europe. (FL learner)

. . . ставит пьесы не только в России, **но и** в Европе

. . . directs shows not only in Russia, **but also** in Europe. (Standard Russian)

(4) человек всегда . . . имеет права не верить \***то, что** общество верет.

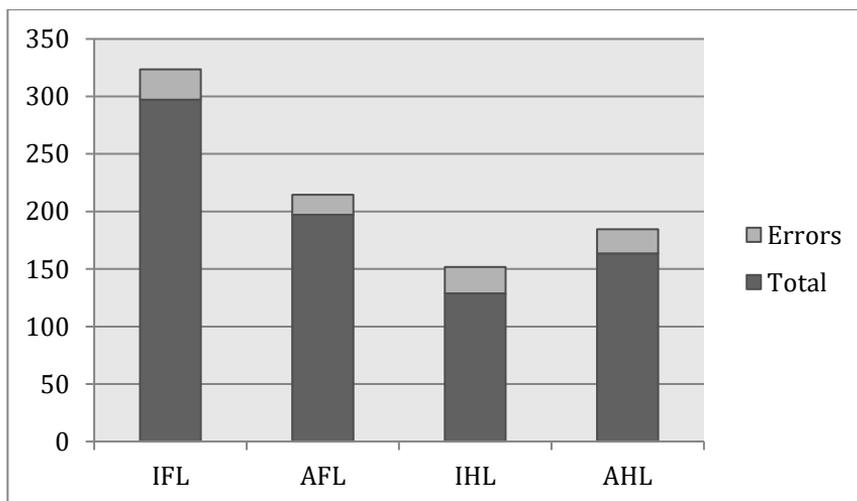
person always . . . has the rights not to believe in that\_ACC, what\_ACC society believes. (FL learner)

человек всегда . . . имеет права не верить **тому, чему** общество верит.

person always . . . has the rights not to believe in that\_DAT, what\_DAT society believes. (Standard Russian)

If we consider the factor of accuracy, there does appear to be a small but discernable development in the HL group. In the Intermediate HL sub-corpus 17% of coordinated sentences contain errors involving coordination and use of conjunction but the percentage of errors goes down to 12.8% in the Advanced HL sub-corpus. However, for the FL writers the percentage of errors in the use of coordinating conjunctions stays stable: 8.9% at the Intermediate level and 8.8% at the Advanced (Figure 2). Most of these errors have to do with the choice between the two adversative conjunctions *a* and *но*, both of which can be translated into English as “but” and are notoriously difficult for American learners of Russian as FL (Dengub and Rojavin, 2010).

Figure 2. Percentage of erroneous sentences with coordination



#### 4.7. Subordinating conjunctions

Subordinating conjunctions are far more numerous in our data. This fact is not surprising since learners at Advanced levels typically engage with higher-level material and tasks that require skills of argumentation, supported opinion, and exposition, which, in turn, require the use of

structures that mark relations between complex ideas. The distribution of frequencies is presented in Table 3.

Table 3. Frequencies of subordinating conjunctions across groups and levels prorated per 1,000 sentences

Subordinating conjunction	Intermediate FL	Intermediate HL	Advanced FL	Advanced HL
что "that"	187.2	140.4	199.1	174.5
почему "why"	2.5	4.6	3.8	2.2
зачем "why"	0	0.4	0	0
как "how, as"	32.2	40.4	31.4	61.3
ли "if"	6.7	10.2	8.8	5
кто "who"	17.1	8.1	11.5	5.6
которые "which"	118.6	83.6	112.1	97.5
после того как "after the fact"	3.9	1.8	1.6	0
до того как "before, up to the point"	1.1	0.4	4	2
с тех пор как "since"	1.1	0.4	0.5	0.3
перед тем как "before the fact"	2.5	0.4	0.8	1.1
как только "as soon as"	0.3	0	0.8	0
когда "when"	38.9	36.5	53	21.3
пока "while"	2	0.4	3.3	1.7
прежде чем "before that"	1.7	0.4	0.8	0.6
если "if"	37.2	27.7	35	26.6
если бы "if, whether"	6.4	3.5	4.6	2
потому что "because"	35.3	23.9	26.8	25.8
так как "because"	1.4	8.8	5.5	14.7
из-за того что "because of"	3.6	0.7	1.4	3.4
в связи с тем что "on the account"	0.6	0.4	0	0.6
в виду того что "for the reason"	1.4	0	0	0

<i>поскольку</i> "since"	2.5	0	6	2.8
<i>ведь</i> "because, after all"	0.8	0.4	1.4	5.9
<i>поэтому</i> "therefore"	18.5	4.9	17.5	7.3
<i>так что</i> "so that"	3.1	0.7	1.4	1.7
<i>тогда</i> "then"	7.6	2.5	3.8	4.5
<i>чтобы</i> "in order to"	62.7	33	56.3	30.8
<i>несмотря на то что</i> "despite"	1.1	1.8	2.7	2.8
<i>хотя</i> "although"	14.3	1.4	23.2	10.9
<i>независимо от того что</i> "regardless"	0.3	0	0	0
<i>чем</i> "than"	11.8	8.1	12.9	9
<i>чем тем</i> "than that"	0.3	0.7	0.8	2
<i>как будто</i> "as if"	1.1	0	1.4	1.1
Total	625.8	446.5	632.2	525

The subordinating conjunctions found in the corpora represent various functional types and reflect various semantic relations between clauses in complex sentences: explanatory, causative, conditional, etc. The subordinating conjunctions were grouped into the following categories:

- (1) explanatory conjunctions and interrogative words in conjunctive function: *что* (cf. that), *кто* (cf. who), *почему* (cf. why), *зачем* (cf. why, to what purpose), *как* (cf. how, that), *ли* (cf. if);
- (2) attributive conjunctions: *какой* (cf. which, that);
- (3) temporal conjunctions: *когда* (cf. when), *после того как* (cf. after), *до того как* (cf. before, up to the point), *с тех пор как* (cf. since), *перед тем как* (cf. while), *пока* (cf. while), etc.;
- (4) conditional conjunctions: *если*, *если бы* (cf. if);
- (5) causal conjunctions: *потому что* (cf. because), *так как* (cf.), *поскольку* (cf. since), *из-за того-*, *в связи с тем-*, *в виду того -* (cf. since, because of), *ведь* (cf. after all);
- (6) conjunctions of result: *поэтому* (cf. therefore), *так что* (cf. so, thus), *тогда* (cf. hence);
- (7) conjunctions of purpose: *чтобы* (cf. in order to);
- (8) comparative conjunctions: *чем* (cf. than), *чем . . . тем* (cf. that . . . than),

как будто (cf. as if); and

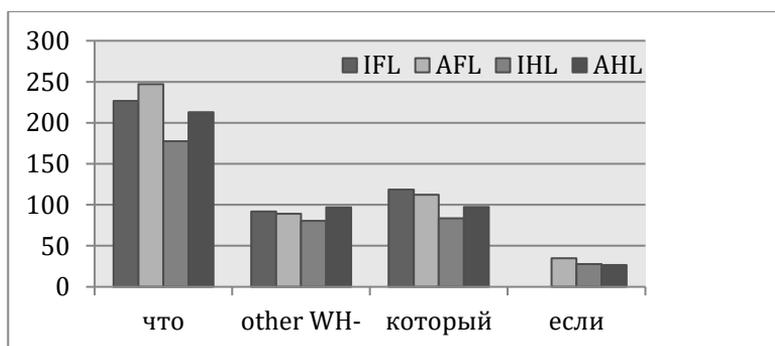
(9) concessive conjunctions: *хотя* (cf. although), *несмотря на то что* (cf. despite the fact that), *независимо от того что* (cf. regardless).

If we consider the most frequent conjunctions in this class, the conjunction *что* (what/that), other interrogative word conjunctions *кто* (who/which), *как* (how/as), *если* (if), and *который* (which), we see a distributional pattern somewhat different from that observed in coordinating conjunctions. The use of *что* (what/that) is on the rise for both groups, with other conjunctions exhibiting relatively static behavior (Table 4, Figure 3).

Table 4. Frequencies of explanatory and attributive conjunctions prorated per 1,000 sentences

	что	other WH-	который	если
Intermediate FL	226.7	91.8	118.6	37.2
Advanced FL	247.2	89	112.1	35
Intermediate HL	177.6	80.3	83.6	27.7
Advanced HL	212.9	97.1	97.5	26.6

Figure 3. Explanatory and attributive conjunctions by group and level



More importantly, the accuracy in the use of explanatory conjunctions is improving for both groups of learners (Figure 4). The most dramatic improvement is observed in the HL groups: at Intermediate level, the HL learners consistently omit punctuation to mark the subordinated clause with *что*, resulting in a 30% error rate in these types

of sentences. At the Advanced level, this rate goes down to 12%. Unlike the HL learners, the FL learners in our study make fewer errors in *что*-coordination (5.5% at the Intermediate level and 3.4% at the Advanced); however, most errors in the FL sub-corpora are structural, signaling the fact that even at the Advanced level, FL learners have difficulties with morpho-syntax. The case of the conjunctive word *который* illustrates the same tendency: the HL learners exhibit difficulties with proper punctuation, whereas the FL learners make errors in agreement.

Figure 4. Errors in the conjunction *что*

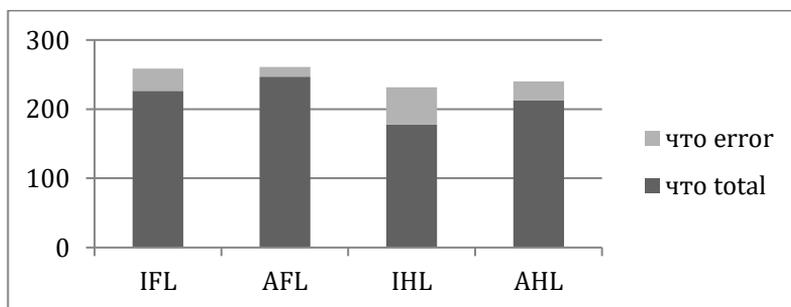
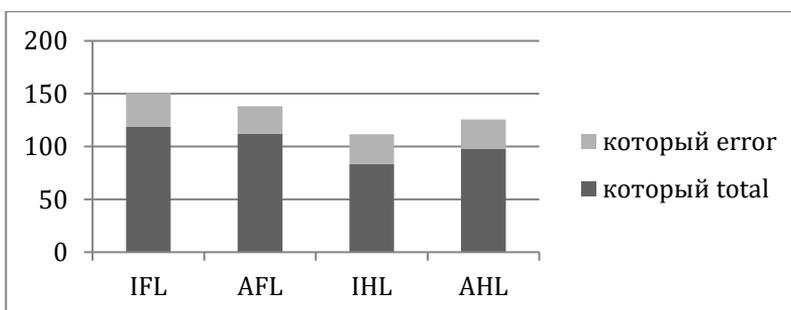


Figure 5. Errors in the conjunctive word *который*



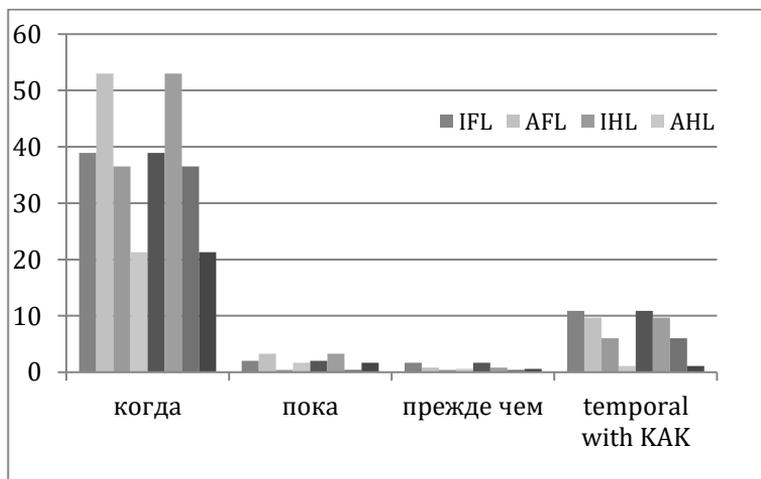
In the temporal conjunctions group, only one conjunction, *пока* “while,” becomes slightly more popular with both groups of learners at Advanced level. The FL group appears to use more *когда* “when” at the Advanced level, while minimizing the use of other temporal conjunctions. It is possible that subordinating clauses with temporal conjunctions are being substituted with phrasal constructions of time (such as during + nominal phrase, e.g., *во время забастовки* “during the protest”); the

transformation of the *когда* clause into a during + nominal phrase structure is structurally relatively easy and may be accessible to the learners. Additionally, both groups of learners make few errors in this category of subordination.

Table 5. Frequencies of temporal conjunctions prorated per 1,000 sentences

	когда	пока	прежде чем	temporal conjunctions with <i>как</i>
Intermediate FL	38.9	2	1.7	10.9
Advanced FL	53	3.3	0.8	9.7
Intermediate HL	36.5	0.4	0.4	6
Advanced HL	21.3	1.7	0.6	1.1

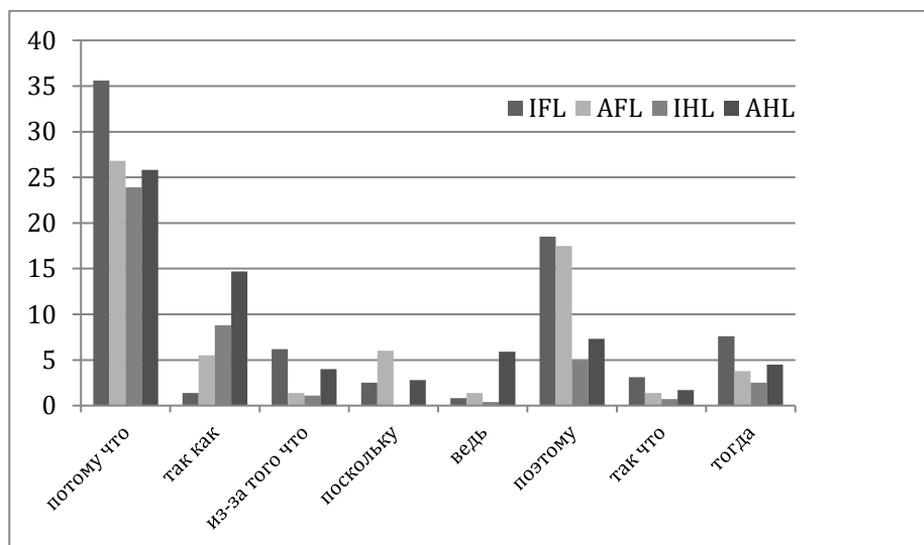
Figure 6. Temporal conjunctions by group and level



Conjunctions of cause, purpose, result, and concession also exhibit an uneven distribution. Interestingly, the use of this functional type increases in the writing of Advanced HL learners, but in our data they still use fewer conjunctions than the FL learners. As with other types of conjunctions, HL learners display improvement in their use of appropriate punctuation. The FL learners mostly produce errors with conjunctions that require structural manipulation with the constituents of

the subordinated clause. In this functional group, the conjunction *чтобы* “in order to” represents such a difficulty for the FL learners, since it requires modifying the verb of the clause to the past tense form. The percentage of these errors, however, is rather small (4% at Intermediate and 2% at Advanced level).

Figure 7. Conjunctions cause, purpose, result, and concession by group and level



In sum, various functional groups of conjunctions and the conjunctions within the groups appear to exhibit uneven distribution, with some discernable patterns of growth or decrease in numbers and some patterns in improvement of accuracy.

## 5. Discussion and conclusion

As a reminder, the questions that guided our study of complex sentences with coordination and subordination were formulated broadly. Approaching the data, we asked: Is there an overall change in the amount, functional variety, and accuracy of complex sentences with conjunctions in the writing of the Advanced learners of Russian as compared to their performance at the Intermediate level? Is there a difference in the trajectory of writing development in regard to complex sentences with conjunctions between the heritage and non-heritage learners of Russian?

The answers to these questions are not straightforward. On the one hand, we observe changes in the numbers of different types of conjunctions in the learner's writing at the two levels of proficiency, but the direction of these changes are different in the FL and the HL groups. The FL learners used fewer coordinated sentences at the Advanced level compared to their performance at the Intermediate level, conforming to our expectations that the more advanced writing or rather more advanced tasks rely more on subordination than coordination. At the same time, the use of complex sentences with coordinating conjunctions by the HL learners increased. This result does not indicate that the HL learners are moving in the opposite direction from the FL learners: we suggest that the developmental trajectories of the two groups of learners are actually converging. Due to different educational histories, FL and HL writers may be simply exhibiting different types of development: for the FL students, this development is more linear and largely shaped by curricular considerations. We can expect the FL learners to move along the curriculum with its tasks first depending more on coordination and then more on subordination. The case of HL learners is different. They usually come to our classes with well-developed oral skills and underdeveloped writing abilities. At the beginning of the course of formal instruction, the HL students have to develop the basics of the written genre and learn to overtly mark the development of thought in the text. Thus, we observe growth in all formal parameters of writing, including overt coordination and overt subordination as discussed below.

The usage of subordinating conjunctions increased for both the FL and HL groups as the learners moved from Intermediate to Advanced levels of proficiency. The increase in subordination was substantial only for the HL group, which utilized 446.5 different subordinating conjunctions per 1,000 sentences at Intermediate level and 525 of these structures at the Advanced level. The FL learners displayed only a minor increase in the use of subordination. Yet, the FL students used more subordinating conjunctions in their writing overall than did their HL peers (632 items in the FL Advanced sub-corpus vs. 525 in the HL Advanced sub-corpus). These observations suggest that subordination continues to play an important role in the creation of logical relationships between parts of a sentence in an Advanced level text and that learners continue to grow their ability to overtly mark logical relationships

between parts of a sentence in their writing.

The dimension of accuracy (i.e., accurate and appropriate use of subordinating structures) appears to play an important role in the development of writing. Accuracy improved for both groups of learners. The HL learners showed the most dramatic increase in the accuracy rates for almost all subordinating conjunctions. However, the overwhelming majority of errors that HL writers committed have to do with missing or misplaced punctuation marks; many teachers would agree that the error gravity of a missing comma (a mechanical error) is lighter compared to morpho-syntactic problems (structural errors) evident in the writing of the FL learners (example 5).

- (5) . . . программы для молодых, \***в которм** можно изучать . . .  
 . . . programs for youth in **which\_PPEP\_masc** one can study . . .  
 (FL learner)  
 . . . программы для молодых, **в которых** можно изучать . . .  
 . . . programs for youth in **which\_PPEP\_pl** one can study . . .  
 (Standard Russian)

To remind the reader, the learners who contributed to the corpus did not receive much formal instruction on the use of complex sentences while in the program. The only explicit feedback they may have received would have been provided in the form of teacher comments and correction on graded writing assignments. On the one hand, this find may suggest that certain mechanical errors (such as punctuation) may improve without targeted and systematic instruction; on the other hand, the FL learners stand to benefit from focus-on-form type of activities, which would target specific conjunctive structures that not only signal logical relations but also require morpho-syntactic modifications (such as the construction with conjunctive word *который* “which” or conjunctions *чтобы* “in order to” and *то, что* “a/the fact that”). Although errors are a necessary stepping-stone in the developmental process, FL learners even at the Advanced level will benefit from targeted instruction on the use of constructions with these conjunctions.

Based on our analysis, we conclude that complex sentences with coordination and subordination especially remain an important linguistic construct, which learners actively utilize at the Advanced level of writing

proficiency. We believe that the number as well as the accuracy and appropriateness of use of coordination and subordination are important indices of proficiency development in the assessment of writing.

At the same time, we recognize the limitation of focusing only on conjunctions. To more fully capture the development in learners' writing ability we must investigate other indices of syntactic complexity as well, such as adverbial and participle clauses, introductory phrases, nominalization patterns, and other structures indicative of syntactic compression, which is believed to be a prominent force in high-level academic writing (Biber et al. 2011). To obtain a more detailed analysis of syntactic complexity, it is crucial to continue to explore the corpus.

Another important issue to consider in future research is the general approach to conjunctions as a grammatical class. The linguistic analysis and classification of conjunctions and their structural and functional characteristics remains a work in progress (Uryson 2017; Zavjalov 2015). Possibly, a more comprehensive, function-based theory on conjunctive devices could lead to a better understanding of potential pedagogical applications. Nonetheless, we find that exploration of a learner corpus provides an invaluable opportunity to conduct research on learner language: the study highlights the segments of the pedagogical grammar that have been operationalized by learners, those which are still emerging, and those requiring substantial curricular and instructional intervention. Even a study as exploratory in nature as the one presented in this paper greatly informed our pedagogical thinking and practice.

## References

- Bardovi-Harlig, Kathleen. 1992. "A Second look at T-unit Analysis: Reconsidering the Sentence." *TESOL Quarterly* 26:390–95.
- Belz, Julia, and Nina Vyatkina. 2008. "The Pedagogical Mediation of a Developmental Learner Corpus for Classroom-Based Language Instruction." *Language Learning and Technology* 12 (3): 33–52.
- Biber, Douglas, Bethany Gray, and Kornwepa Poonpon. 2011. "Should We Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development?" *TESOL Quarterly* 45 (1): 5–35.

- Coe, Richard. 2002. "The New Rhetoric of Genre: Writing Political Briefs." In *Genre in the Classroom: Multiple Perspectives*, edited by Ann M. Johns, 197–207. Mahwah, NJ: Lawrence Erlbaum.
- Dagneaux, Estelle, Sharon Denness, and Sylviane Granger. 1998. "Computer-Aided Error Analysis." *System* 26:163–74.
- Dengub, Evgeny, and Marina Rojavin. 2010. "The Conjunctions и, а, and но: A Comparison of Usage by Heritage and Traditional Students of Russian." *The Slavic and East European Journal* 54:147–64.
- Dengub, Evgeny. 2012. *Investigating Syntactic and Lexical Complexity, Accuracy, and Fluency in the Writing of Heritage Speakers of Russian*. ProQuest LLC.
- Friedman, Debra, and Olga Kagan. 2008. "Academic Writing Proficiency of Russian Heritage Speakers: A Comparative Study." *Heritage Language Education: A New Field Emerging* 181–98.
- Granger, Sylviane. 1996. "From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora." In *Languages in Contrast. Lund Studies in English* 88, edited by K. Aijmer, B. Altenberg, and M. Johansson, 37–51. Lund, Sweden: Lund University Press.
- Granger, Sylviane. 1999. "Use of Tenses by Advanced EFL Learners: Evidence from an Error Tagged Computer Corpus." In *Out of Corpora. Studies in Honour of Stig Johansson*, edited by H. Hasselgard and S. Oksefjell, 191–202. Amsterdam: Rodopi.
- Granger, Sylviane. 2003. "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research." *TESOL Quarterly* 37:538–46.
- Gries, Stefan Th., and Stefanie Wulff. 2013. "The Genitive Alternation in Chinese and German ESL Learners: Towards a Multifactorial Notion of Context in Learner Corpus Research." *International Journal of Corpus Linguistics* 18 (3): 327–56.
- Halliday, Michael A. K. 1985. "Dimensions of Discourse Analysis: Grammar." *Handbook of Discourse Analysis* 2:29–56.
- Hyland, Ken. 2007. "Genre Pedagogy: Language, Literacy and L2 Writing Instruction." *Journal of Second Language Writing* 16 (3): 148–64.
- Leaver, Betty, and Boris Shekhtman. 2002. *Developing Professional-Level Language Proficiency*. Cambridge: Cambridge University Press.

- Leńko-Szymańska, A. (2008). "Non-native or Non-expert? The Use of Connectors in Native and Foreign Language Learners' Texts." *Acquisition et interaction en langue étrangère*, 27: 91-108.
- Lu, Xiaofei. 2011. "A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development." *TESOL Quarterly* 45 (1): 35–62.
- — —. 2012. "The Relationship of Lexical Richness to the Quality of ESL Learners' Oral Narratives." *The Modern Language Journal* 96 (2): 190–208.
- Lu, Xiaofei, and Haiyang Ai. 2015. "Syntactic Complexity in College-Level English Writing: Differences Among Writers with Diverse L1 Backgrounds." *Journal of Second Language Writing* 29:16–27.
- Nesselhauf, Nadja. 2005. *Collocations in a Learner Corpus*. John Benjamins.
- Norris, John M., and Lourdes Ortega. 2009. "Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity." *Applied Linguistics* 30 (4): 555–78.
- O'Donnell, Mathew B., Ute Römer, and Nick C. Ellis. 2013. "The Development of Formulaic Language in First and Second Language Writing: Investigating Effects of Frequency, Association, and Native Norm." *International Journal of Corpus Linguistics* 18 (1): 83–108.
- Ortega, Lourdes. 2003. "Syntactic Complexity Measures and Their Relationship to L2 Proficiency: A Research Synthesis of College-Level L2 Writing." *Applied Linguistics* 24:492–518.
- Osborne, John. 2011. "Fluency, Complexity and Informativeness in Native and Non-native Speech." *International Journal of Corpus Linguistics* 16 (2): 276–98.
- Scott, M. (2014). Research using WordSmith tools.
- Vyatkina, Nina. 2012. "The Development of Second Language Writing Complexity in Groups and Individuals: A Longitudinal Learner Corpus Study." *The Modern Language Journal* 96 (4): 576–598.
- Yang, Weiwei, Xiaofei Lu, and Sara Cushing Weigle. 2015. "Different Topics, Different Discourse: Relationships among Writing Topic, Measures of Syntactic Complexity, and Judgments of Writing Quality." *Journal of Second Language Writing* 28:53–67.
- Wolfe-Quintero, Kate, Shunji Inagaki, and Kim Hae-Young. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, & Complexity*. Honolulu: University of Hawaii Press.

- Барыкина, Алевтина Н., Валентина П. Бурмистрова, Валярия В. Добровольская. 1978. *Пособие по развитию навыков письменной речи*. Москва: Русский язык.
- Завьялов, Виктор Н. 2015. О содержательных составляющих понятия “структура русских союзов”. *Вестник Новосибирского государственного университета. История, филология* 14 (9): 85–90.
- Короткин, Ирина Б. 2011. *Академическое письмо: учебно-методическое пособие для руководителей школ и специалистов образования*. LAP LAMBERT Academic Publishing.
- Лобанова Наталья А., Ирма П. Слесарева. 1980. *Учебник русского языка для иностранных студентов-филологов. Систематизирующий курс*. Москва: Русский язык.
- Ляшевская, Ольга Н., Сергей А. Шаров. 2009. *Частотный словарь современного русского языка*. Москва: Азбуковник.
- Прохорова, Кира В. 1998. *Научный стиль: учебно-методическое пособие для студентов-журналистов*. Санкт Петербург.
- Урынсон, Елена. 2017. Опыт описания семантики союзов. *Лингвистические данные о деятельности сознания*. Litres.
- Химик, В. В. 2003. *Основы научной речи: Учебное пособие для студентов нефилологических высших заведений*. Москва: Академия.