Baltic Journal of English Language, Literature and Culture Vol.2, 2012: 54–62
https://doi.org/10.22364/BJELLC.02.2012.05

SUBORDINATE CLAUSES AS CRITICAL FEATURES IN ENGLISH AND FRENCH LEARNER EXAMINATION CORPORA

VITA KALNBĒRZIŅA AND VINETA RŪTENBERGA
University of Latvia, Latvia

Abstract. The aim of the study is to examine the syntactic pattern of the frequency of use of simple, complex and compound sentences focusing on different subordinate clauses as critical features in the written learner text corpora at different English and French language acquisition levels in the texts produced by secondary school test-takers in Latvia. The theoretical basis of the research is Pienemann's Processability theory. For the purpose of quantitative and contrastive analysis of syntactic structures written learner text corpora in English and French have been compiled, in which different clauses have been marked by manual annotation and afterwards classified according to different syntactic patterns. Subordinate clauses have been grouped applying Dik's taxonomy of embedded constructions. The preliminary research has discovered some similarities and differences between the syntactic pattern usage at the same language acquisition levels.

Key words: learner examination corpora, secondary level, syntactic patterns, subordinate clauses

INTRODUCTION

The necessity of the study is grounded in the popularity of the *Common European Framework of Reference for Languages* (the CEFR) (Online 1), which postulates that all languages are learned in a similar manner, moving from simple short texts to more complex and longer texts and from simple language structures to complex ones. This idea is eagerly adopted by educational administrators across Europe, who demand that language testers produce comparable measurement systems that would function across languages and across age groups. To answer such a demand, it is not enough to use statistical measurements that would show the difficulty levels of the items of each test. What test developers need is reliable indicators that would signal language acquisition levels across languages.

In language testing linguistic corpora enable language testers to better understand the test-takers' language proficiency level. Moreover, they provide evidence of productive language skill level and enhance structural analysis. The best known English language learner corpus is International Corpus of Learner English (ICLE) (Hasselgård, Johanssen, 2011: 37) that contains three million words of essays written by advanced learners of English; the Longman Learners'

Corpus containing ten million words of texts written by students of different proficiency levels and the Cambridge Learner Corpus comprising twenty million words of texts from learners all over the world.

It is considered that the larger the corpora, the more valuable they are. However, for syntactic structure exploration in a learner corpus the size is not as important as for lexical studies, because the number of syntactic variations is rather restricted. Leech (1992) states that 'computer corpora [...] are generally assembled with particular purposes in mind, and are often assembled to be *representative* of some language or text type' (Leech, 1992: 116). It means that a corpus has to be maximally representative of the variables under examination, e.g. different syntactic patterns that are typical of each language acquisition level and could be used as critical features for attributing a certain level to the candidates. According to Sinclair, 'if within the dimensions of a small corpus, using corpus techniques, you can get results that you wish to get, then your methodology is above reproach' (2004: 189). Thus, in order to obtain the most relevant data, the following objectives of the present research were drawn:

a) to compile test-taker written essay corpora in English and French to ensure the representativeness of the learner used syntactic structures across different language acquisition levels;
b) to carry out the empirical investigation of syntactic structures;
c) to provide evidence that syntactic structures can be used as the discriminatory indicators of different language acquisition levels;
d) to compare the syntactic pattern use in English and French at the same language acquisition levels.

LITERATURE REVIEW

The CEFR identifies the linguistic structures that foreign language learners should know at a certain level of language proficiency. For example, at level B1 learners should be able: to understand and produce simple sentences, given that the noun and verb phrases are not overloaded; to understand and produce compound sentences. They are expected to produce complex sentences and be able to understand an embedded clause. At level B2 learners should be able to understand and produce simple, compound and complex sentences.

In Latvia students graduating from a secondary school are to take an examination in one foreign language of their own choice. The candidate performance is to be assessed according to *the same* criteria irrespective of the language. Moreover, language testers have to specify *criterial features* which are defined as 'linguistic properties that are distinctive and characteristic of each of the levels' (English Profile, 2011: 2) validated by empirical research that distinguishes and characterises each of the language acquisition levels, from A to F. In Latvia A is the highest and F – the lowest level of language proficiency. The previous test relation research (Kalnberzina, 2007) suggested that Latvian Year

12 examination level A could be related to the CEFR level C1, Latvian Year 12 levels B and C could be related to the CEFR level B2, levels D and E to levels B1 and A2.

As a theoretical basis for the research, we have chosen Pienemann's approach in analysing linguistic constructions, which is directly linked to the stages in the language acquisition process. He postulates that structural options that may be formally possible will be produced by the language learner only if the necessary processing resources are available (Pienemann, 1999: 2). It means that at a certain stage of development the learner can produce and understand only those linguistic forms which are accessible within human psychology and memory. Pienemann, by applying *processability theory*, shows the order how the main grammatical encoding procedures are activated in syntactic structures in the acquisition of English as a second language. Pienemann suggests that first we acquire a word, then the processes associated with the given word category, then we build phrases based on the categories, develop sentences and add sentence level morphology, e.g. the subject-verb agreement, and at the fifth stage we can build subordinate clauses and use appropriate relationships between the matrix and *subordinate clauses*. Jackson (2007: 54) defines a subordinate clause as 'a clause that does not normally occur on its own, but either in combination with a main clause to form a complex sentence or a part of another clause as an 'embedded' element. [...] Embedded subordinate clauses may function as subject, object or complement in another clause, or as a relative clause.'

It should be stated that Pienemann's hierarchy is implicationally ordered, i.e., every procedure is a necessary prerequisite for the next procedure and it reflects the time-course in language generation, which could be relied on when analysing syntactic patterns at different language acquisition levels. Therefore, it was decided to focus on syntax as it facilitates the understanding of how the process of communication and interaction among humans develops, how sentences are constructed because sentence structure expresses the most important grammatical relationships in all human languages.

METHODS

The present research is a corpus-based quantitative and contrastive analysis of subordinate clauses in English and French learner examination corpora. McEnery et al. consider corpus-based studies 'as a methodology with a wide range of applications across many areas and theories of linguistics' (2006: 9). Biber, Conrad and Reppen (1998) state the main reasons why corpus-based studies have become more common nowadays:

- they are empirical, analysing the actual patterns of use in natural texts;
- they utilize a large and principled collection of natural texts, known as a 'corpus', as the basis for analysis;

- they make extensive use of computers for analysis, using both automatic and interactive techniques;
- they depend on both quantitative and qualitative analytical techniques (Biber et al., 1998: 4).

The quantitative analysis supplies information on *frequency* of different subordinate clauses (noun, adjectival and adverbial). Frequency 'in the text is the instantiation of probability in the system. A linguistic system is inherently probabilistic in nature. [...] to interpret language in probabilistic terms, the grammar [...] has to be able to represent language as *choice,* since probability is the probability of 'choosing'' (Halliday, 2005: 45). Moreover, having the frequency information from a corpus, we can 'establish the probability profile of any grammatical system' (ibid.: 67). The frequency is subdivided into three groups:

1) 'Raw frequency' is simply a count of how many instances of some linguistic phenomenon X occur in some corpus, text or collection of texts;
2) 'Normalized frequency' (sometimes called 'relative frequency') expresses frequency relative to a standard yardstick (e.g. 'tokens per million words');
3) 'Ordinal frequency', the frequency of X is compared with the frequencies of Y, of Z, etc. (Leech, 2011: 7-8).

When speaking about the learner corpora, *ordinal frequency* is the most important measure to be used. This is why Year 12 exam of the English and French language written examination corpora, consisting of essays, have been developed and annotated to produce empirically measurable results that are not predictable only from language learning theories. Moreover, the contrastive analysis of the obtained data has been used as it is of utmost importance to prove the assumption that *all* languages are learned in a similar manner, moving from simple short texts to more complex and longer texts, as stated by the CEFR.

PROCEDURE

The compiled English learner examination corpus consists of 44387 words; while the French learner examination corpus contains 28378 words (see Appendices 1, 2). When developing the corpus, written texts were chosen from the year 2009 centralised examination of secondary school graduates to represent all language acquisition levels, both in English and French. However, it has to be stated that texts of levels E and F in French could not be obtained as the number of test-takers per year is approximately 120 (in 2011 it was even less – 77 students) and the examination results range mainly from levels A to D. It should be specified that as the aim of the contrastive analysis is to examine syntactic patterns across different languages, we do not focus on discourse analysis in this study.

In English the test-takers had to write an essay about *'Reasons for Leaving Latvia'*:

> One of the main reasons why people left Latvia during the last few years is that they say they are better paid in other countries. Add two other reasons and discuss all of them in an argumentative essay, giving your own opinion.

The following was the theme of the essay in French:

> Pensez vous qu'il soit encore utile d'apprendre des langues étrangères alors que l'anglais est actuellement la langue de communication mondiale (échanges commerciaux, économiques, politiques...)? Présentez votre réflection de façon argumentée. (Do you think that it is still useful to learn foreign languages as nowadays English is the language of communication (in business, economics, politics…) in the world? Give your point of view by providing arguments.)

In the corpora different sentences (simple, compound and complex) were marked by manual annotation. Afterwards the complex sentences were in the focus and they were classified by using Dik's (1997) taxonomy of embedded constructions. The taxonomy could be attributed across languages, which is of major importance as there exist different ways of producing various embedded constructions. Dik distinguishes between finite and non-finite embedded constructions. All main clauses contain a finite verb, but it is not the case with all embedded clauses. They might have a non-finite verb, hence the division of the embedded constructions into finite and non-finite. According to Dik (1997: 144), finite embedded constructions are 'those [...] in which the predicate can be specified for the distinctions which are also characteristic of main clause predicates.' This is very obvious in cases when a subordinate clause may appear also as an independent main clause. Another important aspect that subordinate clauses are marked by subordinating devices was taken into consideration, too. Thus, the subordinate clauses were divided into noun, adjectival and adverbial clauses. It should be noted that the first subordination that follows directly the matrix clause was chosen as the clause discriminating element.

RESULTS AND DISCUSSIONS

The research results show that simple sentences have been used extensively as they appear at all levels of language proficiency both in English and French (in English their number ranges from 19 at level B to 5 at levels D and E while in French – from 10 at level B to 3 at level C). The same cannot be said about the compound sentences as the test-takers of English and French have used a limited number of compound sentences in their written production. In English the number of compound sentences at level A is only 2, at level B it reaches 15 sentences and then at levels C and D the number falls again to 2 at level D. At levels E and F no compound sentences have been produced. In French the numbers range from 2 sentences at level A to 4 sentences at level B and then fall down to only 1 sentence at level D. When examing the complex sentences, we see that their number increases at the higher levels of language proficiency in English. Figure 1 shows that the number of complex sentences reaches 26 at level A, which is almost three times more than at level B. At level C it rises to 13, but at lower levels D-F their number is only 4 to 5. In French the number of complex sentences at level A reaches 5 sentences, which is lower than at level B where their number is 10 (Figure 2). If compared with English, their number is lower at the highest level of language proficiency (level A), but higher at level D. Thus, the sentence distribution is not the same in both languages.

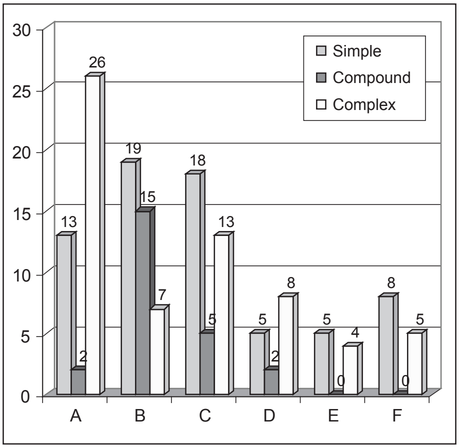


Figure 1 **Distribution of sentences in the English texts**

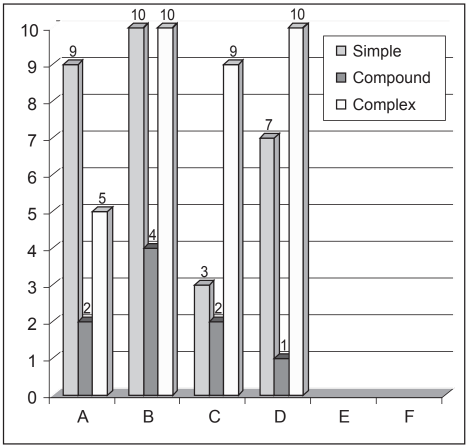


Figure 2 **Distribution of sentences in the French texts**

The complex sentences were chosen for further analysis and they were classified into three groups (noun, adjectival and adverbial clause) according to the first subordination which follows directly the main clause. Figure 3 shows that in English the number of noun clauses is surprisingly high only at level A reaching the number 11, while all the other levels comprise one, two or three cases. In French (Figure 4) the highest number of noun clauses appears at level D, which is 4, while at the higher levels A and B the number varies from 1 to 2 clauses. As for adjectival clauses, the test-takers of English have used them more at level A (6 times) while in French their number is rather limited at all difficulty levels (1 or 2). Adverbial clauses have been used most frequently both in English and French, and they predominate at all levels. In English they range from 9 at levels A and C to 4 at level D, 3 at levels B, F and 2 at level E. In French adverbial clauses reach the number 7 at level B and then fall to 5 at level C, 4 at level D and 3 at level A. According to the obtained data the adjectival clauses appear to be the most discriminating as their number compared with the other clauses is rather limited.

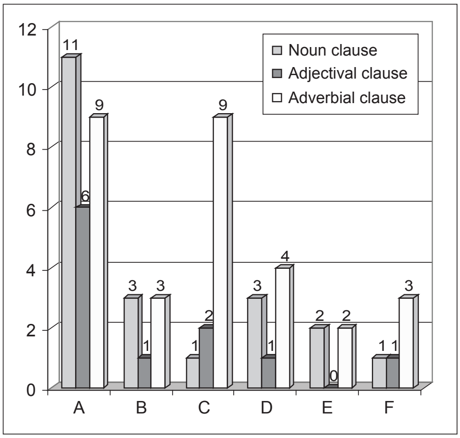


Figure 3 **Frequency of subordinate clauses in the English texts**

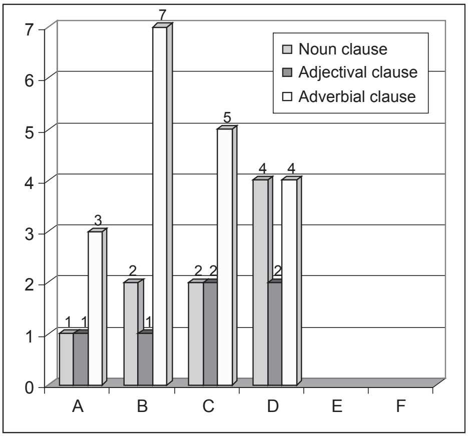


Figure 4 **Frequency of subordinate clauses in the French texts**

The results of contrastive analysis prove the assumption based on Pienemann's Processability theory that syntax is one of the parameters signalling a certain level of language acquisition, and that subordinate clauses serve as a criterial feature for attributing higher levels at the language examination. However, the obtained data of subordinate clauses differ in English and French which might signal problems in the development and interpretation of the writing assessment criteria and standardisation of markers. Therefore, further research of the frequency of noun, adjectival and adverbial clauses at different language acquisition levels both in English and French is of utmost importance.

CONCLUSIONS

The preliminary comparison of the clause profiles in the English and French written texts across the levels suggests the following:

1. at levels D and E the learners rely more on adverbial and noun clauses both in English and French;
2. at levels B and C the learners in both languages use more adverbial clauses;
3. at level A the learners of English use rather many adjectival clauses, but still noun and adverbial clauses predominate, while the learners of French use few noun and adjectival clauses at level A.

The present research has examined the role of learner corpora in test equation across the languages and has discovered some similarities and differences between the syntactic patterns used at the same language acquisition levels. Further research is necessary to examine the reasons why the number of adjectival and adverbial clauses differs – whether the variability is caused by the small size of the corpora and the lack of representativeness of the sample, the differences in the test tasks, the learner proficiency levels, or whether it is caused by the differences in syntactic patterns in the English and French languages.

REFERENCES

Biber D., Conrad S., Reppen R. (1998) *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Dik, S.C. (1997b) *The Theory of Functional Grammar, Part 2: Complex and Derived Constructions.* K. Hengeveld (ed.). Berlin: Mouton de Gruyter.

English Profile. Introducing the CEFR for English, (2011) Cambridge: Cambridge University Press

Halliday, M.A.K. (2005) *Computational and Quantitative Studies.* J.J.Webster (ed). London and New York: Continuum.

Hasselgård, H., Johanssen, S. (2011) Learner Corpora and Contrastive Interlanguage Analysis. In F. Meunier, S. De Cock, G. Gilquin and M. Paquot (eds). *A Taste for Corpora : In honour of Sylviane Granger.* Amsterdam and Philadelphia: John Benjamins.

Jackson, H. (2007) *Key terms in linguistics*. New York: Continuum.

Kalnberzina ,V. (2007) *Impact of Relation of Year 12 English Language Examination to CEFR on the Year 12 Writing Test.* International Conference of the FIPLV Nordic-Baltic Region 'Innovations in Language Teaching and Learning in the Multicultural Context'. Riga, Latvia.

Leech, G. (1992) Corpora and theories of linguistic performance. In J. Svartvik (ed) *Directions in Corpus Linguistics* (pp 105-22). Berlin: Mouton de Gruyter.

Leech, G. (2011) Frequency, corpora and language learning. In F. Meunier, S. De Cock, G. Gilquin and M. Paquot (eds). *A Taste for Corpora : In honour of Sylviane Granger.* Amsterdam and Philadelphia: John Benjamins.

McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies. An Advanced Resource Book.* London and New York: Routledge.

Pienemann, M. (1999) *Language Processing and Second Language Development: Processability Theory.* Amsterdam/Philadelphia: John Benjamins.

Sinclair, J. (2004) *Trust the Text: Language, Corpus and Discourse.* New York: Routledge.

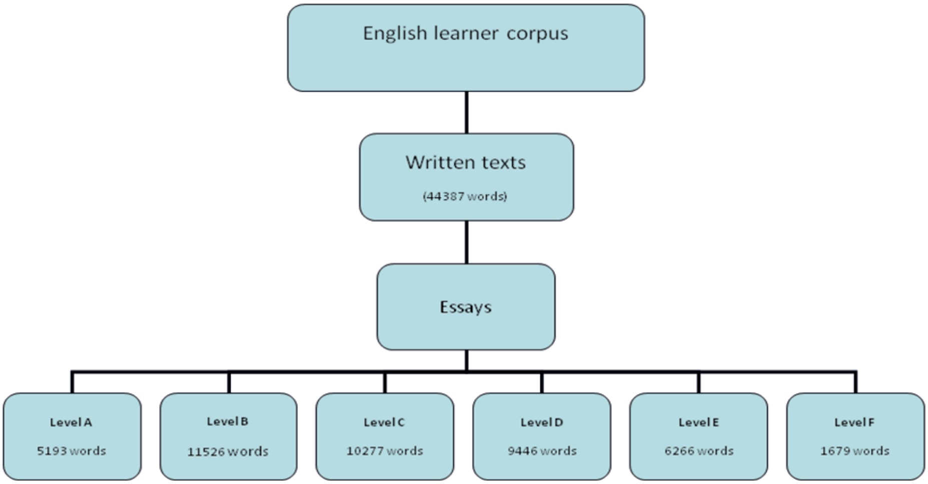
Svartvik, J. (1973) *Errata: Papers in Error Analysis*. Lund: Gleerup/Liber.

INTERNET SOURCES

1) Available from *http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf* [Accessed September 5, 2011].

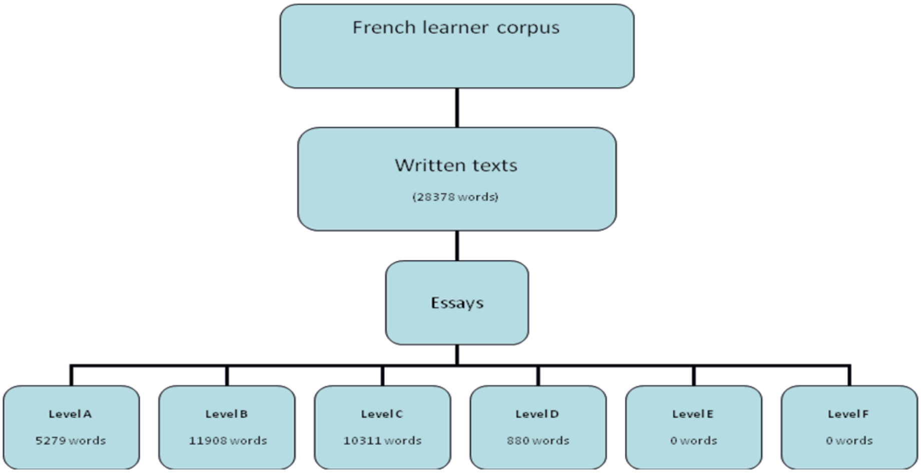
APPENDIX 1

Year 12 English learner written corpus consisting of 44387 words.



APPENDIX 2

Year 12 French learner written corpus consisting of 28378 words.



Vita Kalnbērziņa (Dr. Phil., Assoc. Prof.) is currently working at the University of Latvia. Her research interests include language testing and language acquisition. Email: *vita_kalnberzina@yahoo.com*

Vineta Rūtenberga (MA Philol.) is currently doing her PhD studies on the 'Utilisation of the Universal Grammar in Assessing Different Language Essays' at the University of Latvia. Her research interests are foreign language learning and testing. Email: *vrutenberga@hotmail.com*