Vineta Rūtenberga
Latvijas Universitāte, Latvija

SYNTACTIC STRUCTURES – A DICRIMINATORY FACTOR OF FOREIGN LANGUAGE PROFICIENCY LEVEL

Abstract
The Common European Framework claims that all languages are learned in a similar way, starting from simple phrases and moving to more complex sentences and longer texts. Therefore, language testers need to have reliable indicators that would signal language proficiency levels across languages.
Thus, different syntactic structures in the written learner text corpora at different English and French proficiency levels in the texts produced by the secondary school graduates were examined to provide evidence that syntax is a discriminatory indicator of language proficiency levels.
The empirical part of the research is based on the quantitative and contrastive analysis of simple, compound and complex sentences in the written learner text corpora, mostly focusing on descriptive statistics.
The theoretical basis of this study is Pienemann's *Processability theory* (1999), which postulates 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. It provides the order how the main grammatical encoding procedures are activated in syntactic structures in the acquisition of English.

Key words: language acquisition, test-taker corpus, syntactic structures

There has been a remarkable interest in language acquisition over the centuries, but the study of how people acquire a foreign language has prospered at the end of the previous century with the necessity to communicate, obtain education or compete in the job market. Nowadays the interest in foreign language acquisition has not diminished as there is even a higher demand for mobility, international communication, information access, mutual understanding, etc.

Chomsky maintains that language acquisition is an activity unique to human beings, and different in kind from any other type of learning which human beings experience (Radford 1998:8). Moreover, it is an activity that all human beings possess irrespective of their intelligence level. Children are born with a so-called language acquisition device, which is known also as Universal Grammar. Although the language input that children get from their parents may be grammatically incomplete, they are able to solve these problems within a certain period of time by producing grammatically correct language.

Cognitive psychologists who investigate the information processing model of human learning and performance "tend to see second language acquisition as the building up of knowledge systems that can eventually be called on automatically for speaking and understanding" (Lightbown 1999: 41).

One of the most recent studies on second language acquisition has revealed that language learners also undergo sequences of development. Pienemann elaborated the "processability theory" as so far learnability was considered as purely logico-mathematical problem.

Thus, we see that there are different approaches and attempts to explain the processes involved in language acquisition as it is a very complicated mechanism. The study of a language depends on which aspect of language the linguists are interested in, whether they want to find out what is common to all languages; how sentences are structured grammatically or what words, sentences and texts mean. Recently, different corpora have been used for the previously mentioned study purposes as the actual language is studied in the texts occurring naturally.

Corpus linguistics is rather a new approach to language, which emerged at the same time when Chomsky focused on the theory of syntax, i.e. in the 1960s. The description of various languages, which were considered to be a universal phenomenon, was not satisfactory. Certain grammatical features typical for one particular language were insufficiently described. Chomsky was not interested in language beyond the sentence level. For him authentic data were not of any significance as grammar was considered to be autonomous, independent of meaning. Sinclair, on the contrary, argued that "language should be studied in naturally occurring contexts of use – and should have at its centre the analysis of meaning" (2004: 2). He postulated that lexical and syntactic patters could not be separated, as well as competence in a language could not be separated from performance in the use of that particular language. Therefore, real language data for more precise empirical analysis were required.

The first large-scale language data collection for empirical grammatical research was held by Randolph Quirk in the late 1950s. The first modern corpus of the English language known as the Brown corpus was created in 1960s. But only in the mid-1980s with the advances in technology, the obtained data were computerized. It facilitated the further studies of language, because computerized corpora can be processed very rapidly, accurately and the result is more reliable. Besides the corpus texts can be used for different linguistic analyses.

Nowadays a corpus can be defined as "a body of naturally occurring language" (McEnery et al. 2006: 4). Sinclair (1996) stressed that "a corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language." There have been different definitions of what a corpus is, but there are two features that the linguists agree upon, namely, a corpus is: 1) machine-readable; 2) contains authentic texts.

Nowadays corpus-based studies have become more common because of several reasons stated by Biber, Conrad and Reppen (1998: 4):

- it is empirical, analysing the actual patterns of use in natural texts;
- it utilizes a large and principled collection of natural texts, known as a "corpus", as the basis for analysis;

- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques.

Corpus compilation, namely test-taker written essays, implies creating electronic version of the target texts, which is rather labour intensive and time consuming process as they have to be keyboarded manually. A simple corpus could contain only target texts, without any additional information about the authors, structure, contents.

The most important thing is to define what it is that the particular corpus represents. Only afterwards, the frequency of the particular feature is to be investigated as the central focus is on repeated elements rather than on single occurrences.

Thus, corpus studies can provide valuable resources for those interested in the particular research. Further on, I will focus on how syntactic structures are produced in acquiring a language as this is the aspect to be researched in the test-taker written corpora.

There are different approaches to defining what *syntax* is. Chomsky considered syntax to be the central aspect of language. He defines syntax as the study of the principles and processes by which sentences are constructed in particular languages (2002: 11). In his *Principle and Parameters* approach to syntax Chomsky states that in every human language there is a set of universal principles which are known by all human beings. In addition to universal principles, there are a finite number of parameters which define how to apply the universal principles to construct grammatical sentences. Thus, in generative grammar sentences are generated by a subconscious set of procedures.

Liddicoat (2007) states that syntax deals with how to put words together to form sentences which mean what we want. Rather similar definition is given by Tallerman where syntax means “sentence construction”: how words group together to make phrases and sentences (2005: 1).

In all the mentioned definitions the main focus of syntax is how to construct sentences because sentence structure expresses the most important grammatical relationships in all human languages.

Pienemann postulated that structural options that may be formally possible will be produced by the language learner only if the necessary processing resources are available (1998: 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 (1 - lemma access; 2 - category procedure; 3 - phrasal

procedure; 4 - S-procedure/WO Rules; 5 - subordinate clause procedure). There is a time sequence involved in producing grammatical constructions, e.g., noun and verb phrases appear before sentences; category procedure appears before phrase procedure, etc. Moreover, he points out that "Universal Grammar has been productive mostly as a property theory, addressing the issue of the origin of linguistic knowledge (i.e. the "logical problem") and has been far less successful in accounting for the "developmental problem"" (ibid.: 34).

The basic reasons (VanPatten, 2008: 141) why language learners follow this hierarchy are:

- the hierarchy is implicationally ordered, that is, every procedure is a necessary prerequisite for the next procedure;
- the hierarchy mirrors the time-course in language generation.

With reference to the previously mentioned reasons, this hierarchy was also chosen as the basis for a more profound study of different sentences, in particular the complex ones, as they appear at the very top of the hierarchy, which means that subordination is the highest level of language proficiency.

In the exploratory research I have examined: 1) the frequency of use of simple, compound and complex sentences; 2) different types of subordinate clauses in complex sentences in the written learner text corpora at different English and French language acquisition levels.

The texts were produced by test-takers graduating from the secondary school and having studied English and French from 3 to 12 years. For the exploratory research twelve texts from English and twelve texts from French written learner corpora (two from each level A - F) were chosen (it should be stated that it was rather difficult to find texts of levels E and F in French as the number of test-takers per year comprises approximately 120 and they are mainly pupils from language schools).

First, different syntactic structures were marked by hand in the chosen test-taker texts. Afterwards, the obtained data were classified according to the level (A-F) obtained at the centralised state exam in English and French. Furthermore, the analysis of subordinate clauses was carried out.

The obtained results show (see *Figures 1, 2*) that at the lowest levels of language proficiency (E, F) mainly simple sentences predominate, though in French complex sentences are more frequent at levels B-E than at level A. Simple and complex sentences occur alternatively. This might be explained either by the fact that the test-taker texts are assessed subjectively or the criteria according to which the written performance is assessed do not discriminate well between different language proficiency levels. Overall, it is evident that syntactic structures serve as one of the indicators of the language acquisition levels, especially in English.

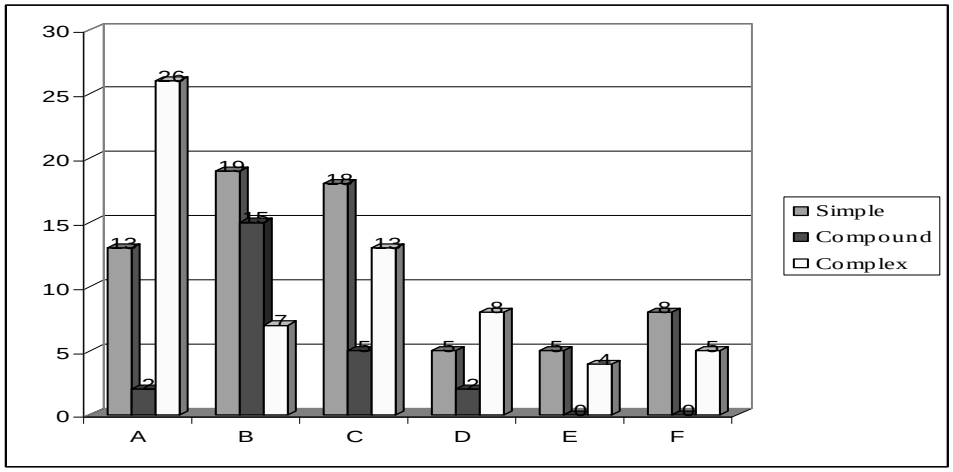

Figure.1 **Distribution of sentences in the English texts**

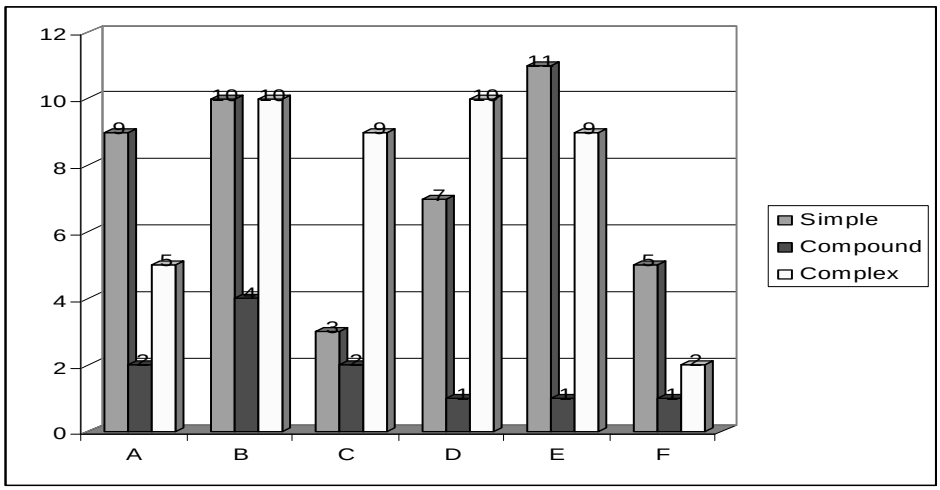

Figure 2 **Distribution of sentences in the French texts**

Furthermore, the analysis of subordinate clauses was carried out as according to Pienemann's "Processability theory" as well as the secondary language examination specifications the use of subordinate clauses is one of the main indicators of the higher level of linguistic competence.

The complex sentences were classified in three groups according to the subordinate clause division. As the examples showed that the majority of complex sentences contained several subordinate clauses, the first subordination that followed directly the matrix clause was chosen as a clause discriminating element.

Adjectival clauses are considered to be the most complicated of complex sentence subordinate clauses, which is also evident from the obtained data analysis.

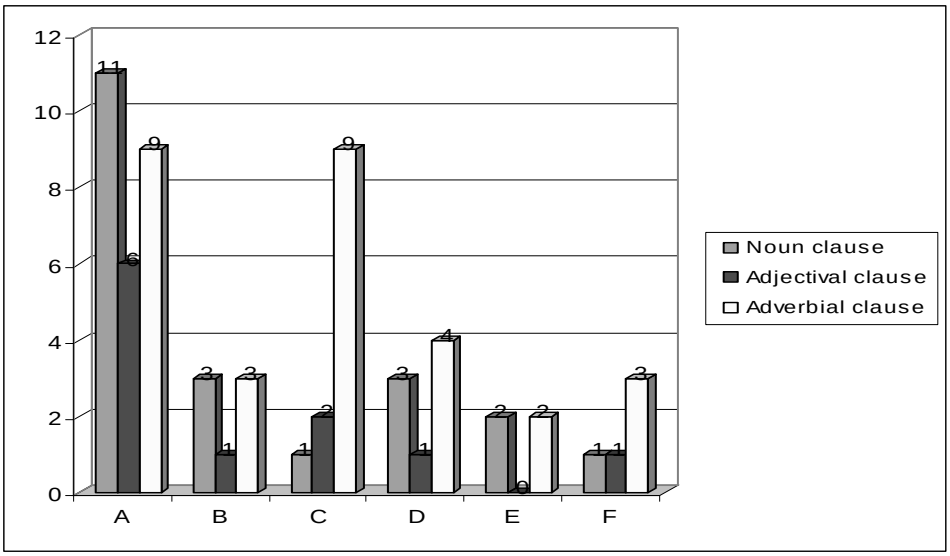

Figure 3 **Frequency of subordinate clauses in the English texts**

All three subordinations appear at the highest level of the English test-taker texts and the adjectival clause rises above the others at level A.

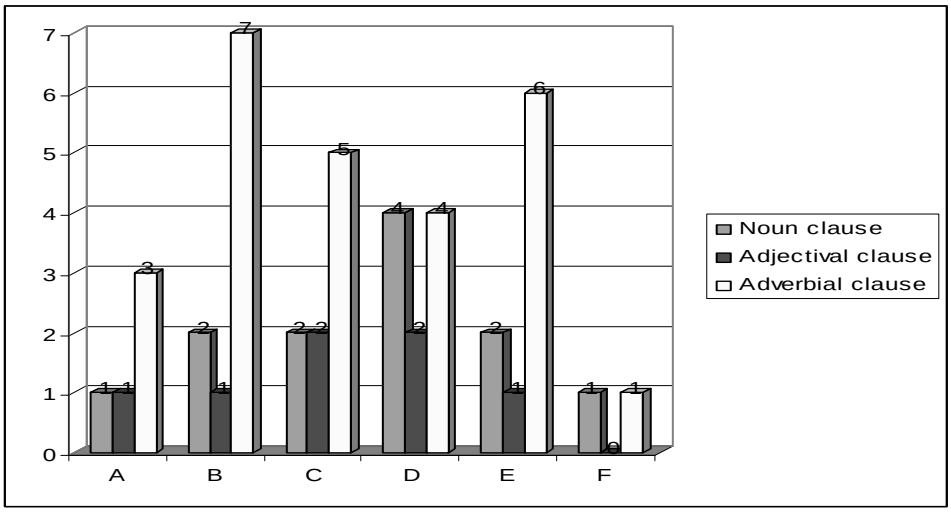

Figure 4 **Frequency of subordinate clauses in the French texts**

The situation is not similar in French. It is obvious that there is a limited number of noun and adjectival clauses in the French texts but the adverbial clauses predominate in all levels of language proficiency.

The exploratory research validates the application of the above mentioned theories in analysing the main indicators in different language proficiency levels. The data demonstrate that there is a certain order in learning various syntactic patterns, the highest of which being the adjectival clause, and that the subordinate clause serves as an indicator of the test-taker language proficiency level. However, the contrastive analysis of English and French do not

provide the same data for both languages, e.g., in French the distribution of clauses according to the obtained level do not serve as a discriminating factor. Thus, in order to provide a more reliable evidence of sentence structures which indicate the difference among language proficiency levels, a deeper analysis of a larger corpus is required.

Bibliography

1. Chomsky, N. (2002) *Syntactic Structures*. Second Edition. With an Introduction by David W.Lightfoot. Berlin, New York: Mouton de Gruyer.
2. Crystal, D. (2007) *How language works*. London: Penguin books.
3. Lightbown, P.M., Spada, N. (1999) *How Languages are Learned*. Second edition. Oxford University Press.
4. McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies: and advanced resource book.* New York: Routledge.
5. Pienemann, M. (1999) *Language Processing and Second Language Development: Processability Theory*. Studies in bilingualism 15. Amsterdam/Philadelphia: John Benjamins Publishing Company.
6. Radford, A. (1998) *Syntax: a minimalist introduction*. Cambridge University Press.
7. Reppen and Simpson (2002) Corpus Linguistics. In Schmitt, N. (ed.), *Introduction to Applied Linguistics.*(pp.92-111) Great Britain: Arnold.
8. Sinclair, J. (2004) *Trust the Text: language, corpus and discourse.* New York: Routledge.
9. Tallerman, M. (2005) *Understanding Syntax*. Second edition. Malta: Hodder Arnold.
10. VanPattern, B., Williams, J. (ed.), (2008) *Theories in Second Language.* New York: Routledge.