

BABEȘ-BOLYAI UNIVERSITY
CLUJ-NAPOCA
FACULTY OF LETTERS

THE LEXICAL INVENTORY OF ROMANIAN AS A
FOREIGN LANGUAGE. LEVEL A1

PhD Thesis
SUMMARY

SCIENTIFIC COORDINATOR:

EMERITUS PROF. DR.: GAVRILĂ NEAMȚ

PhD candidate: Antonela-Carmen Suciu (married Arieșan-Simion)

2020

CONTENTS

ARGUMENT	5
SECTION I - THEORETICAL FRAMEWORK.....	10
Chapter I – The role of the European linguistic documents in aligning Romanian as a foreign language to international standards	11
1.1. The Common European Framework of Reference for Languages: learning, teaching, assessment	11
1.2. The CEFR Companion Volume with New Descriptors	25
1.3. Materials developed based on CEFR	28
1.4. Lexical competence.....	30
1.5. Conclusions	39
Chapter II – Observations regarding the present stage of the research.....	40
2.1. The case of other international languages	41
2.1.1. The case of English	41
2.1.2. Profile Deutsch and the German lexical inventory	46
2.1.3. The case of Spanish.....	47
2.1.4. Conclusions	49
2.2. The present stage of the research domain in Romania.....	49
2.2.1. A practical dictionary for foreign students. 1000 Romanian words.....	50
2.2.2. The vocabulary of minimal Romanian for foreign students	51
2.2.3. A Minimal Romanian dictionary with grammar instructions and translation into French, English and Spanish	52
2.2.4. Antonyms. Synonyms. Analogies. The vocabulary of minimal Romanian (with English translation).....	52
2.2.5. Threshold Level in learning Romanian as a foreign language.....	53
2.2.6. A minimal description of Romanian as a foreign language. A1, A2, B1, B2	54
2.2.7. An attempt at level assessment.....	55
2.2.8. The Romanian main lexical fund	56
2.2.9. The representative vocabulary of Romance languages	58
2.2.10. Romanian Frequency List	59
Chapter III – On the necessity and relevance of lexical inventories	63
3.1. On the necessity of lexical ranking in RFL.....	65
3.2. CEFR perspective on level-based lexical ranking.....	71

3.3. Lexical inventory vs. dictionary	72
3.4. The utility of lexical inventories in the teaching/learning/assessment process.	74
3.5. Conclusions	77
Chapter IV – Corpus linguistics, in general	79
4.1. Corpus linguistics. Classifications	82
4.2. CoRoLa – Computational Corpus of Reference for Romanian as a foreign language ..	83
4.3. ROMBAC – The Romanian Balanced Annotated Corpus	84
4.4. RoCo_News	84
4.5. The advantages of a written productions corpus of non-native speakers.....	86
4.6. Corpus-based linguistics. Interdisciplinary applicability	96
4.7. Conclusions	98
Chapter V – Romanian vocabulary. General framework	99
5.1. Romanian vocabulary. General considerations	99
5.2. Perspectives in approaching lexical knowledge in L2	102
5.3. Principles to be followed in generating ranked lists of words	105
5.4. Criteria.....	110
5.4.1. Range.....	111
5.4.2. Frequency	111
5.4.3. Dispersion.....	112
5.4.4. Lexical sets criterion	113
5.4.5. Efficiency	115
5.4.6. The subjective criterion.....	116
5.4.7. The combinatorial capacity	116
5.5. The delimitation of word number for each level of linguistic competence.....	117
5.6. Conclusions	119
Chapter VI – The role of statistics in corpus-based linguistics	121
6.1. Statistics and LC.....	121
6.2. The corpus of written productions of Romanian as a foreign language non-native speakers. Level A1	129
6.3. The description of the written productions of non-native speakers, level A1	131
6.3.1 The subjects.....	132
6.3.2. The productions.....	135
6.3.4. Transcription productions	138
6.4. CORLS – Oral corpus of Romanian as a foreign language	139

6.5. Corpus of assignment requests	141
6.6. Corpus of Romanian as a maternal/native language	142
6.7. Conclusions	142
SECTION II – RESEARCH RESULTS	143
Chapter VII – Research results.....	144
7.1. Methodology	144
7.2. Quantitative analysis	145
7.3. Proper nouns.....	156
7.4. Acronyms and abbreviations	162
7.5. Numbers	165
7.6. Exclusions from the frequency list.....	166
7.6.1. Marks of oral discourse	166
7.6.2. Transfers from other languages.....	168
7.6.3. Special characters or letters.....	169
7.7. Qualitative analysis	169
7.7.1. Collocations specific to level A1	170
7.8. Lexical spheres completion criterion	179
7.9. Necessity criterion.....	182
7.10. Zipf’s law applied to the frequency list of the written production corpus of RFL non- native speakers	186
7.11. Conclusions	188
CONCLUSIONS.....	196
BIBLIOGRAPHY	201
ANNEXES	216
Annex I – Written corpus of Romanian as a foreign language, Level A1	217
Annex II – Frequency list, final version.....	431
Annex III – Frequency list based on the requirements of the teaching-learning materials...	450

Key-words: native language (L1), second language (L2), foreign language (FL), target language (TL), input, output, level of linguistic competence, level A1, corpus, corpus of non-native speaker productions, microlanguage, lexical inventory, list of words, word, lexeme, sense, corpus-based linguistics, linguistics statistics, absolute frequency (AF), relative frequency (RF), range (R).

A synthesis of the main parts of the doctoral thesis:

The purpose of this research is to approach a theme which has gained international interest lately, but did not gain the same attention in the Romanian research domain, namely the elaboration of an A1 level lexical description, outlined as lexical inventory for Romanian as a foreign language, according to the *Common European Framework of Reference for languages: learning, teaching, assessment* (CEFR).

The necessity and the relevance of a lexical ranking of Romanian as FL/L2 was mentioned on many occasions by specialists, such as: Elena Platon, Gabriela Biriş, Gabriela Stoica, Diana Burlacu, Adriana Todea sau Carmen Mîrzea Vasile. Such an instrument would generate a great advantage to the RFL domain, as it could foster a higher degree of unity in the way the A1 level vocabulary is understood by specialists. Moreover, from an educational and instructional point of view, it becomes a concrete support in the elaboration of curricula and teaching, learning and assessment materials, while from a scientific perspective, it represents a point of reference for future research in the domain of Romanian as L1, L2, as well as other comparative studies.

The corpus of written productions of non-native RFL speakers became the main objective of the present research, being an element of novelty for the RFL domain, as up until now, in accordance with present knowledge, no one has published such a type of corpus for Romanian as a foreign language (except for the oral productions, CORLS). In the present analysis, the corpus was exploited using quantitative methods (FA, FR, R calculus), and qualitative ones (the concordances), in order to ensure illustrative results on what A1 level means from a lexical point of view, for non-native speakers of Romanian as FL/L2.

In the present approach, an important role is played by reference works, such as: *Common European Framework of Reference for languages: learning, teaching, assessment* (CEFR), *The CEFR Companion Volume with New Descriptors, Reference Level Descriptions – RLD^s for national and regional languages*, *The threshold level in learning Romanian as a foreign language*, *The minimal description of Romanian. A1, A2, B1, B2*, *The minimal vocabulary of Romanian for foreign students*, *Making and Using Word Lists for Language*

Learning and Testing, Statistics in Corpus Linguistics. A practical Guide and Essential Word List.

As presented in the Contents, the present paper consists of two main *sections*, preceded by an *Argument* and followed by general *Conclusions, Bibliography* and three *annexes*.

The first section addresses the theoretical aspects and consists of six chapters, briefly presented in what follows:

Chapter 1, *The role of the European linguistic documents in aligning Romanian as a foreign language to international standards*, addresses European reference works with regards to linguistic policies (CEFR, Companion Volume, Threshold Level), discussing and analysing, in general terms, concepts and scales with proposed descriptors, with special focus on lexical concepts.

Chapter 2, *Observations regarding the present stage of the research*, begins with a presentation of the context for English, Spanish and German, highlighting the strengths and the limitations of each described material. The second part of this chapter presents the internal aspects discussing works that analyse the lexicon taught to non-natives (*The Threshold Level in learning Romanian as a foreign language, A minimal description of Romanian as a foreign language. A1, A2, B1, B2, A practical dictionary for foreign students. 1000 Romanian words, The vocabulary of minimal Romanian for foreign students, Antonyms. Synonyms. Analogies. The vocabulary of minimal Romanian (with English translation), A Minimal Romanian dictionary with grammar instructions and translation into French, English and Spanish, Romanian Frequency List*), as well as papers that support such claims (*The Romanian main lexical fund, or The representative vocabulary*).

Chapter 3, *On the necessity and relevance of lexical inventories*, is a chapter that displays the advantages of a lexical inventory, from an instructional perspective, in the domain of RFL. It also presents the CEFR perspective on this topic and underlines the main differences between inventories and dictionaries.

Chapter 4, *Corpus linguistics, in general*, presents, as stated in the title, the corpus linguistics, with a presentation of what a corpus is, of the linguistic corpora approaches and classifications, followed by a presentation of the main electronic corpora of Romanian as a native language (ROMBAC; CoRoLa, RoCO_news). This chapter also addresses the corpora on non-native speakers, underlining the advantages both from a scientific and instructional perspective, as well as the concept of *microlanguage*, as defined by Elena Platon, within the Romanian research framework.

Chapter 5, *Romanian vocabulary. General framework*, opens up with a presentation of the Romanian vocabulary and a review of the main approaches to lexical learning. The second part of this chapter describes and analyses the principles used in the Western research framework in elaborating lists of words, as well as the statistical and subjective criteria taken into consideration in the case of this type of analysis.

Chapter 6, *The role of statistics in corpus-based linguistics (LC)*, focuses on the presentation of the main methods and statistical concepts used in corpus-based linguistics, as part of the field of statistics, in order to submit and interpret the results of the analysis. In this chapter, the written corpus of the RFL non-native speakers, level A1 is being presented. It addresses target group aspects, main topics, collection modalities and transcription conventions. Last but not least, the three corpora used in the analysis are briefly presented: CORLS, ROMBAC and a small corpus which includes the assignments and the requirements of the materials used in the teaching and assessment of RFL.

The second section presents the results of the analysis based on the corpora of written and oral productions of RFL non-native speakers, level A1, and concludes with a presentation of the lexical inventory of Romanian as a foreign language, for level A1, as well as the proposal of an additional list, meant as a companion for this linguistic instrument (proper nouns list, acronyms and abbreviations, together with the list of memorized formulae). The present analysis highlights what the non-native speaker can do and understand, insisting upon what should be contained in the microlanguage, from a lexical point of view, at level A1. For this purpose, alongside the statistical criteria, some criteria with subjective character have also been used – the criteria of necessity, lexical sets and subjectivity. Finally, the lexical inventory of the Romanian language as FL/L2, level A1, contains a number of 839 words and 120 memorized formulae.

In conclusion, the present research benefits, on one hand, the domain of Romanian as a foreign/second language, and on the other, that of Romanian as a maternal language.