

UNIVERSITATEA BABEȘ-BOLYAI
CLUJ-NAPOCA
FACULTATEA DE LITERE

INVENTARUL LEXICAL AL LIMBII ROMÂNE
CA LIMBĂ STRĂINĂ.
NIVEL A1
Teză de doctorat
REZUMAT

COORDONATOR ȘTIINȚIFIC:

PROF. UNIV. DR. EMERIT: GAVRILĂ NEAMȚ

Doctorand: Antonela-Carmen Suciu (căs. Arieșan-Simion)

2020

CUPRINS

ARGUMENT	5
SECȚIUNEA I - CADRU TEORETIC	10
Capitolul I – Rolul documentelor lingvistice europene în procesul de aliniere a limbii române ca limbă străină la standardele internaționale.....	11
1.1. Cadrul European comun de referință pentru limbi: învățare, predare evaluare	11
1.2. The CEFR Companion Volume with New Descriptors.....	25
1.3. Materiale elaborate pornind de la CECR	28
1.4. Competența lexicală.....	30
1.5. Concluzii	39
Capitolul II – Observații asupra stadiului actual al cercetării.....	40
2.1. Situația altor limbi de circulație internațională.....	41
2.1.1. Cazul limbii engleze	41
2.1.2. Profile Deutsch și inventarul lexical al limbii germane.....	46
2.1.3. Situația limbii spaniole	47
2.1.4. Concluzii.....	49
2.2. Situația actuală a cercetării pe teren românesc	49
2.2.1. Dicționar practic pentru studenți străini. 1000 de cuvinte românești	50
2.2.2. Vocabularul minimal al limbii române pentru studenți străini	51
2.2.3. Dicționarul minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă	52
2.2.4. Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză	52
2.2.5. Nivel Prag pentru învățarea limbii române ca limbă străină.....	53
2.2.6. Descrierea minimală a limbii române ca limbă străină. A1, A2, B1, B2.....	54
2.2.7. Încercare de plasare pe nivel.....	55
2.2.8. Fondul principal lexical al limbii române.....	56
2.2.9. Vocabularul reprezentativ al limbilor romanice	58

2.2.10. Romanian Frequency List	59
Capitolul III – Despre necesitatea și relevanța inventarelor lexicale.....	63
3.1. Despre necesitatea ierarhizării lexicului RLS	65
3.2. Perspectiva CECR asupra lexicului ierarhizat pe niveluri	71
3.3. Inventar lexical vs. Dicționar	72
3.4. Utilitatea inventarelor lexicale în procesele de predare/învățare/evaluare	74
3.5. Concluzii	77
Capitolul IV – Corpusul lingvistic, în general	79
4.1. Corpusul lingvistic. Clasificări	82
4.2. CoRoLa – Corpus Computațional de Referință pentru limba română ca limbă contemporană	83
4.3. ROMBAC – The Romanian Balanced Annotated Corpus	84
4.4. RoCo_News	84
4.5. Avantajele unui Corpus de producții scrise ale vorbitorilor nonnativi	86
4.6. Lingvistica bazată pe corpus. Aplicabilitate interdisciplinară	96
4.7. Concluzii	98
Capitolul V – Vocabularul limbii române. Cadru general	99
5.1. Vocabularul limbii române. Considerații generale	99
5.2. Perspective în abordarea cunoașterii lexicale în L2.....	102
5.3. Principii de urmat în elaborarea listelor de cuvinte ierarhizate	105
5.4. Criterii	110
5.4.1. Răspândirea.....	111
5.4.2. Frecvența.....	111
5.4.3. Dispersia	112
5.4.4. Criteriul întregirii sferelor lexicale	113
5.4.5. Eficiența	115
5.4.6. Criteriul subiectiv.....	116
5.4.7. Capacitatea combinatorie.....	116
5.5. Delimitarea numărului de cuvinte pe fiecare nivel de competență lingvistică	117

5.6. Concluzii	119
Capitolul VI – Rolul statisticii în lingvistica bazată pe corpus (LC)	121
6.1. Statistica și LC	121
6.2. Corpusul de producții scrise ale vorbitorilor nonnativi de limba română ca limbă străină. Nivel A1	129
6.3. Descrierea corpusului de producții scrise ale vorbitorilor nonnativi, nivel A1	131
6.3.1 Subiecții	132
6.3.2. Producțiile	135
6.3.4. Convenții de transcriere	138
6.4. CORLS - Corpus oral de limba română ca limbă străină	139
6.5. Corpusul de cerințe	141
6.6. Corpusul de limba română ca limbă maternă	142
6.7. Concluzii	142
SECȚIUNEA A II-A – REZULTATELE CERCETĂRII	143
Capitolul VII – Rezultatele cercetării	144
7.1. Metodologie	144
7.2. Analiza cantitativă	145
7.3. Substantivele proprii	156
7.4. Acronime și abrevieri.....	162
7.5. Numerele.....	165
7.6. Excluderi din lista de frecvențe.....	166
7.6.1. Mărci ale discursului oral.....	166
7.6.2. Transferuri din alte limbi	168
7.6.3. Caractere speciale sau litere.....	169
7.7. Analiza calitativă	169
7.7.1. Colocații specifice nivelului A1	170
7.8. Criteriul întregirii sferelor lexicale	179

7.9. Criteriul necesității.....	182
7.10. Legea lui Zipf aplicată listei de frecvențe din corpusul de producții scrise ale vorbitorilor nonnativi de RLS	186
7.11. Concluzii	188
CONCLUZII.....	196
BIBLIOGRAFIE.....	201
ANEXE	216
Anexa I - Corpus scris de limba română ca limbă străină, Nivel A1	217
Anexa II – Lista de frecvență, versiunea finală	431
Anexa III – Lista de frecvență din cerințele materialelor de predare-evaluare.....	450

Cuvinte-cheie: limbă maternă (L1), limbă a doua (L2), limbă străină (LS), limbă-țintă (LȚ), competență, input, output, nivel de competență lingvistică, nivelul A1, corpus, corpus de producții ale vorbitorilor nonnativi, microlimbă, inventar lexical, listă de cuvinte, cuvânt, lexem, sens, lingvistică bazată pe corpus, statistică lingvistică, frecvență absolută (FA), frecvență relativă (FR), răspândire (R).

Sinteza părților principale ale tezei de doctorat:

Prin această cercetare ne-am propus să abordăm o temă de actualitate în plan internațional, însă care pe teren românesc nu a beneficiat încă de suficientă atenție, elaborarea unei descrieri lexicale a nivelului A1, în forma unui inventar lexical al limbii române ca limbă străină, conform *Cadrului european comun de referință pentru limbi: învățare, predare, evaluare* (CECR).

Necesitatea și relevanța unei ierarhizării a lexicului limbii române ca LS/L2 a fost evidențiată în mai multe ocazii, de specialiști precum: Elena Platon, Gabriela Biriș, Gabriela Stoica, Diana Burlacu, Adriana Todea sau Carmen Mîrzea Vasile. Marele avantaj pe care acest instrument l-ar aduce domeniului RLS este reprezentat de oferirea unui grad mult mai ridicat de unitate în ceea ce înțelege fiecare specialist prin lexicul de nivel A1. În plus, în plan didactic, constituie un suport concret pentru elaborarea curriculelor și a materialelor de predare, învățare și evaluare, iar din perspectivă științifică, reprezintă un reper pentru cercetările viitoare din domeniul limbii române ca L1, L2, precum și pentru alte studii comparative.

Pentru a ne atinge obiectivul cercetării noastre, am elaborat un corpus de producții scrise ale vorbitorilor nonnativi de RLS, nivel A1, care constituie un alt element de noutate pentru domeniul RLS, întrucât la ora actuală, după cunoștințele noastre, nu există disponibil un alt corpus de acest fel elaborat pentru limba română ca limbă străină (exceptându-l pe cel de producții orale, CORLS). În analiza noastră am exploatat corpusul folosind atât metode cantitative (calculul FA, FR, R), cât și calitative (concordanțele), pentru a ne asigura de obținerea unor rezultate ilustrative a ceea ce înseamnă nivelul A1 din punct de vedere lexical pentru un vorbitor nonnativ de română ca LS/L2.

În demersul nostru, un rol important l-au avut lucrări de referință, precum: *Cadrul european comun de referință pentru limbi*, *The CEFR Companion Volume with New Descriptors*, *Reference Level Descriptions – RLD^s for national and regional languages*, *Nivel Prag pentru învățarea limbii române ca limbă străină*, *Descrierea minimală a limbii române. A1, A2, B1, B2*,

Vocabularul minimal al limbii române pentru studenți străini, Making and Using Word Lists for Language Learning and Testing, Statistics in Corpus Linguistics. A practical Guide și Essential Word List.

Așa cum rezultă din cuprins, lucrarea noastră este formată din două *secțiuni* principale, precedate de un *Argument* și urmate de o parte de *Concluzii* generale, *Bibliografie* și de trei *anexe*.

Prima secțiune, care este destinată părții teoretice a lucrării, este formată dintr-un număr de șase capitole, prezentate pe scurt, în continuare:

Capitolul 1, *Rolul documentelor lingvistice europene în procesul de aliniere a limbii române ca limbă străină la standardele internaționale*, tratează lucrările de referință, în plan european în ceea ce privește politicile lingvistice (CECR, Companion Volume, Nivel Prag), discutând și analizând conceptele și scalele cu descriptori propuse, în general, și pe cele cu privire la lexic, în special.

Capitolul 2, *Observații asupra stadiului actual al cercetării*, începe cu o prezentare a situației pentru limbile engleză, spaniolă și germană, evidențiind punctele tari, dar și limitele fiecărui material descris. În cea de-a doua parte a acestui capitol prezentăm situația în plan intern luând în discuție lucrări care tratează lexicul predat nonnativilor (*Nivel Prag pentru învățarea limbii române ca limbă străină, Descrierea minimală a limbii române. A1, A2, B1, B2, Dicționar practic pentru studenți străini. 1000 de cuvinte românești, Vocabularul minimal al limbii române pentru studenți străini, Antonime. Sinonime. Analogii. Vocabular minimal al limbii române (cu traducere în limba engleză), Dicționarul minimal al limbii române cu indicații gramaticale uzuale și traducere în franceză, engleză, spaniolă, Romanian Frequency List*), dar și lucrări despre care s-ar putea susține acest lucru (*Fondul principal lexical al limbii române sau Vocabularul reprezentativ*).

Capitolul 3, *Despre necesitatea și relevanța inventarelor lexicale*, este un capitol în care evidențiem avantajele pe care un inventar lexical le aduce, în special, în plan didactic domeniului RLS. Tot aici, arătăm care este perspectiva CECR asupra acestei teme și subliniem principalele diferențe între inventare și dicționare.

Capitolul 4, *Corpusul lingvistic, în general*, tratează, după cum o spune și titlul, corpusul lingvistic, făcând o prezentare a ceea ce presupune un corpus, abordări și clasificări ale corpusurilor lingvistice, după care sunt prezentate principalele corpusuri electronice de limba română ca limbă maternă (ROMBAC; CoRoLa, RoCO_news). Tot în cadrul acestui capitol

discutăm și despre corpusurile vorbitorilor nonnativi, subliniind avantajele pe care acestea le aduc atât în plan științific, cât și didactic, precum și despre conceptul de *microlimbă* așa cum apare la Elena Platon, pe teren românesc.

Capitolul 5, *Vocabularul limbii române. Cadru general*, se deschide cu o prezentare a vocabularului limbii române și cu o trecere în revistă a principalelor abordări în învățarea lexicală. În a doua parte a acestui capitol, am descris și analizat principiile folosite în spațiul anglofon în elaborarea listelor de cuvinte, precum și criteriile statistice, precum și pe cele subiective de care se ține cont în cazul acestui tip de analiză.

Capitolul 6, *Rolul statisticii în lingvistica bazată pe corpus (LC)*, se axează pe prezentarea principalelor metode și concepte statistice care sunt utilizate de lingvistica bazată pe corpus din aria statisticii pentru a prezenta și a interpreta rezultatele analizelor. Tot în cadrul acestui capitol, prezentăm corpusul scris al vorbitorilor nonnativi de RLS, nivel A1. Discutăm despre grupul-țintă, temele abordate, modalitatea de colectare sau convențiile de transcriere. Nu în ultimul rând, prezentăm succint alte trei corpusuri folosite în analiză: CORLS, ROMBAC și un mic corpus format din cerințele unor materiale de predare-evaluare RLS.

A doua secțiune prezintă rezultatele analizei pe baza corpusurilor de producții scrise și orale ale vorbitorilor nonnativi de RLS, nivel A1, și se încheie prin prezentarea inventarului lexical al limbii române ca limbă străină, pentru nivelul A1, și a unor liste suplimentare, propuse a însoți acest instrument lingvistic (lista substantivelor proprii, a acronimelor și a abrevierilor, împreună cu lista formulelor memorate). În analiza noastră, am încercat să evidențiem ceea ce poate face și înțelege vorbitorul nonnativ, arătând ceea ce ar trebui să conțină microlimba din punct de vedere lexical la nivelul A1. În acest scop, pe lângă criteriile statistice, am folosit și unele cu caracter subiectiv – criteriul necesității, al întregirii sferelor lexicale sau al subiectivității. În final, inventarul lexical al limbii române ca LS/L2, nivel A1, conține un număr de 839 de cuvinte și 120 de formule memorate.

În concluzie, considerăm că prin cercetarea noastră aducem un beneficiu atât domeniului limbii române ca limbă străină/limbă a doua, cât și limbii române ca limbă maternă.