VOCABULAR MEDICAL PENTRU LIMBA ROMÂNĂ CA L2 DIN PERSPECTIVA CORPUSURILOR ELECTRONICE

GABRIELA BIRIȘ

Facultatea de Litere, Universitatea din București

gabriela.biris@litere.unibuc.ro

Cuvinte-cheie: corpus medical specializat, corpus lexical academic, analiză lexicală, limba română ca L2.

Keywords: specialized medical corpus, academic lexical corpus, lexical analysis; Romanian as a second language (L2).

1. INTRODUCERE

Articolul propune o analiză a conținuturilor lexicale din corpusurile digitale medicale și din cursurile academice românești publicate după 2000, destinate vorbitorilor străini, cu accent pe tipologia și frecvența cuvintelor și pe contextele în care sunt utilizați termenii medicali, în vederea obținerii unor criterii de categorizare cu scop didactic a inventarelor lexicale medicale.

Dezvoltarea mijloacelor digitale de procesare a limbajului natural a condus, în ultimele decenii, la elaborarea unui număr mare de corpusuri electronice, cu diferite funcționalități: antrenarea și evaluarea modelelor de limbaj, traducere automată, stilometrie, analize de frecvență a cuvintelor, modele statistice de discurs etc. Aceste corpusuri oferă date empirice referitoare la frecvența, comportamentul colocațional sau gramatical al cuvintelor etc., utile și domeniilor lingvistice[1].

Pentru predarea limbajelor de specialitate la nivel pre-medical (An pregătitor), corpusurile specializate au un conținut prea complex și dimensiuni foarte mari spre a fi utilizate ca atare, nefiind corpusuri academice, dar pot contribui la evidențierea caracteristicilor generale ale textelor biomedicale, utile în redactarea materialelor didactice pentru limba română ca L2. De exemplu, corpusurile medicale specializate românești de texte scrise sincronice și autentice[2] au reliefat următoarele:

[1] Pentru o prezentare a corpusurilor din perspectivă lingvistică, vezi Carmen Mîrzea-Vasile 2017, Elena Faur și Andreea-Nora Pop 2017.

[2] ROMBAC include prospecte farmaceutice (Ion et al,. 2012), Corpusul românesc biomedical din cadrul proiectului CoRoLa acoperă nouă subdomenii medicale (neurologie, diabetologie,

SCL, LXXVII, 2026, nr. 2, Bucureşti, p. 163–172

DOI: 10.59277/SCL.2026.2.02

– distribuția părților de vorbire: 60% substantive, 18% adjective, 17.6% verbe și 4.4% adverbe (Ion *et al.* 2012: 344),

– densitatea lexicală a textelor: o medie de 20.28 de cuvinte pe frază (Mitrofan 2017),

– raportul de 1:3 între cuvintele purtătoare de semnificație și cele gramaticale (Ion *et al.* 2012: 343).

Se confirmă astfel și pentru limba română rezultatele înregistrate în studiile anglo-saxone, pentru domeniul medical: caracterul nominal și descriptiv al textelor, frecvențele mari ale cuvintelor gramaticale, densitatea lexicală ridicată și complexitatea sintactică (Salager-Meyer 1987, 1983, 1992, Ferguson 2001). Gruparea conceptelor în patru tipuri semantice, în corpusul SiMoNERo: anatomie (ANAT), substanțe chimice și medicamente (CHEM), boli și afecțiuni (DISO) și proceduri medicale (PROC)(Mitrofan și Păiș 2022) se suprapune parțial peste categorizarea cu care s-a lucrat în proiecte anterioare (anii ’70-’80), care au vizat limbajul medical de la anul pregătitor[3]: biologie umană, chimie organică, farmacologie. Cu precizarea că ceea ce s-a lucrat înainte de ’90 a avut la bază criteriul frecvenței termenilor în texte scrise (manuale de liceu), prelucrate manual (Marinescu, Petre *et al.* 1987: 180).

Schema de adnotare utilizată în corpusul digital menționat, cu următoarele tipuri semantice: structură anatomică, localizare și regiune din corp, parte a corpului, organ sau componentă a organului, zonă a corpului sau joncțiune, substanță a corpului, sistem, celulă, componentă celulară, structură embrionară, structuri anatomice formate și țesut, preluată de la National Library of Medicine[4], confirmă partiția propusă pentru sistematizarea tematică a vocabularului medical destinat Anului pregătitor (Șoșa *et al.* 1990, Biriș 2015a).

Conținuturile lucrărilor academice de limbaj medical adresate vorbitorilor străini urmează două orientări: a. centrarea pe informații de anatomie (cu subteme: alcătuirea corpului uman, niveluri de organizare, structuri, aparate, sisteme, elemente de fiziologie), precum în Mișcalencu *et al.* 1983, Biriș 2015 b, Dafinoiu 2023 și b. acoperirea unor informații mai generale, cu caracter de popularizare științifică medicală, ca în Velican-Aldea 1991, Bălănescu 1998 și Argăseală *et al.* 2022, ambele orientări fiind susținute de texte adaptate.

Abordarea vocabularelor specializate prin mijloace digitale se află încă la început în domeniul limbii române ca L2, un domeniu de cercetare pe care l-a explorat și lingvista omagiată astăzi, prof. dr. Oana Murăruș, prin contribuții relevante[5].

---

endocrinologie, cardiologie, oncologie, nefrologie, neurologie, psihiatrie, chirurgie și medicină alternativă)(Mitrofan și Tufiș 2018), iar SiMoNERo (sau MoNERo) conține texte din literatura științifică medicală, articole și bloguri (Mitrofan și Păiș 2022: 317).

[3] Pentru o prezentare detaliată a acestor cercetări, Biriș 2015a.

[4] https://www.nlm.nih.gov/research/umls/knowledge_sources/semantic_network/index.html.

[5] *Teste de limba română: Pașaportul lingvistic European*, în colaborare cu Andra Vasilescu și Camelia Ușurelu, București: Editura Nouă, 2006; *Limba română ca limbă străină: Dosare pedagogice*, în colaborare cu Liviu Groza, Dana Niculescu și Andra Vasilescu (coord.), București: Editura Universității din București, 2006.

2. DATELE DIN CORPUSUL ELECTRONIC

Corpusul SiMoNERo[6] oferă acces liber la inventarul de intrări/tokenuri (engl. *tokens*), la datele privind frecvența acestora și la contextele în care au fost identificate. Este un corpus de tip arbore sintactic (engl. *syntactic treebank*), cu 146.020 de tokenuri și peste 4200 de fraze, adnotat sintactic și morfologic. Accentul s-a pus în acest corpus pe domeniile: cardiologie, diabetologie și endocrinologie (Mitrofan și Păiș 2022: 317). Cele mai numeroase ocurențe înregistrează cuvintele gramaticale, de exemplu *de* (6724), *și* (3749), *în* (3629), *la* (2556), verbele *a fi* (3090) și *a avea* (2212). Între intrările cu indice de frecvență mai mare sau egal cu 100, adică primele 200 din lista de leme, apar opt termeni anatomici: *celulă, cardiac, renal, vascular, sânge, cardiovascular, arteră* și *genă,* cinci termeni medicali de uz general: *pacient, tratament, boală, bolnav, sarcină,* precum și cinci termeni medicali specializați: *clinic, chirurgical, cronic, sindrom, disfuncție* și *insuficiență*.

Într-un eșantion format din 1000 de tokenuri, am identificat trei categorii de termeni medicali: anatomici (abreviat TA), de uz general (UG) și specializați (TS), eliminând termenii din patru categorii: boli și afecțiuni (de exemplu, *infarct, AVC, hipertensiune, disfuncție* etc.), substanțe (de exemplu, *glucoză, insulină* etc.), proceduri medicale (de exemplu, *bypass*, *postoperatoriu* etc.), precum și pe cei care țin de specializări și practică medicală (de exemplu, *chirurgical, cronic* etc.), deoarece ar fi depășit sfera de interes a unui curs pre-medical. În cazul familiilor lexicale, cum ar fi: *genă – genetic, pacient – pacientă, sânge – sanguin* etc., a fost inclus un singur termen, cel cu frecvența mai mare. A rezultat un inventar de 69 de termeni (Anexa 1), cu următoarea pondere rotunjită:

– 41% termeni medicali specializați;

– 33% termeni anatomici;

– 26% termeni medicali generali.

Categorizarea semantică s-a făcut din perspectivă didactică, favorizând uzul, cu mențiunea că un termen poate glisa între categorii, inclusiv înspre cele excluse menționate anterior, sau poate fi distribuit simultan în două categorii, în funcție de context. Cu toate că „anatomic” și „specializat” nu se exclud, am adoptat dicotomia deoarece există o diferență de nivel de specializare între termeni din perspectiva unei L2 Putem vorbi despre un nivel „de bază” pentru *craniu, membru, abdomen, nucleu* etc., pe care l-am considerat anatomic propriu-zis, și un nivel și un nivel cu grad de dificultate ridicat, pentru *sinusal, fetal* sau *miocardic*, la care vorbim despre complexitate lingvistică (opacitate semantică, derivare, compunere, etimon complex etc.) și conceptuală, deci „specializat”.

[6] Consultat între 19 iulie și 29 august 2026, la https://github.com/UniversalDependencies/UD_Romanian-SiMoNERo. Prin bunăvoința colegelor Elena Irimia și Verginica Barbu Mititelu de la Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu” al Academiei Române, cărora le adresez întreaga mea gratitudine, am beneficiat de explicații tehnice cu privire la acest corpus. Exprim mulțumiri speciale Elenei Irimia pentru generarea inventarului de leme în format Excel.

Structura eșantionului indică o pondere apropiată între termenii anatomici și cei specializați, precum și frecvențe mari la termenii de uz general: *boală* (335), *pacient* (1034), *tratament* (453), *terapie* (190). Implicația didactică ar consta în necesitatea unei selecții lexicale graduale, care să pornească de la termenii medicali generali și anatomici și să avanseze către terminologia specializată. Se confirmă rolul vocabularului medical general ca fundament al comprehensiunii, în timp ce ponderea termenilor anatomici și specializați implică necesitatea introducerii progresive.

3. DATELE DIN CURSURI ACADEMICE

În vederea alcătuirii inventarului lexical al cursurilor academice de limbaj medical, am utilizat programul gratuit AntConc[7] pentru analize lingvistice de text și corpusuri. Programul a permis generarea listelor de tokenuri după criteriul alfabetic și după cel al frecvenței, identificarea asocierilor lexicale (engl. *collocations*), a grupurilor de cuvinte (engl. *clusters*) și a contextelor anumitor cuvinte (engl. *KWIC, key word in context*), precum și compararea celor trei corpusuri rezultate, câte unul pentru fiecare dintre cursurile studiate: Biriș 2015 b, Argăseală *et. al.* 2022 și Dafinoiu 2023. Inventarele generate electronic au fost analizate calitativ și adnotate manual, pentru a unifica formele flexionare (de exemplu, *mușchii, mușchiului, mușchilor* au fost incluse sub *mușchi*), și pentru a categoriza semantic și în funcție de uz terminologia medicală identificată.

Și în aceste cursuri, ponderea cea mai mare o au cuvintele gramaticale (*de, și, în, la*), împreună cu verbele *a fi* și *a avea*, ambele categorii cu ocurențe de ordinul sutelor, dar și o serie de pronume (demonstrative: *acesta*, relative: *care*) sau diverse numerale, aspect care ar putea confirma proporția de 1 la 3 stabilită pentru SiMoNERo. Totodată, numeroase ocurențe înregistrează cuvintele de metalimbaj didactic, care apar în structurarea conținuturilor și în formularea aplicațiilor: *exercițiu, exemplu, completa, corecta, enunț, întrebare, răspuns, spațiu, vocabular* etc.

Din perspectivă cantitativă, corpusurile (denumite după primul autor) sunt apropiate ca număr de tokenuri: Biriș – 5631, Argăseală *et al.* – 5443, Dafinoiu – 5042. Ca și în cazul analizei pentru SiMoNERo, am lucrat pe eșantioane de 1000 de tokenuri, ordonate descrescător după frecvență, pentru fiecare curs în parte, respectiv pe un corpus unificat cu 11. 116 tokenuri, pentru evidențierea aspectelor comune celor trei cursuri.

În corpusul Biriș, datorită structurării tematice a unităților de curs (aparate și sisteme anatomice, cu elemente minime de fiziologie), s-au putut observa ocurențe numeroase la anumite verbe relaționate descrierilor de poziție, structură sau componentelor corpului uman: *a cuprinde* (128), *a se găsi* (141), *a se afla*

[7] Disponibil la https://www.laurenceanthony.net/software/antconc.

(157), *a prezenta* (121), *a conține* (95), *a exista* (89), *a alcătui* (72), *a forma* (58). Dintr-un total de 5631 de tokenuri, pe eșantionul cu care am lucrat, am obținut o listă cu 166 de termeni medicali, având următoarea pondere: termeni anatomici 45%, termeni de uz general 14% și termeni specializați 41%. Terminologia medicală generală înregistrează și aici ocurențe numeroase: *medic, pacient, boală, sănătate, tratament, durere* și *medicament*. Diferențierea între termenii anatomici propriu-ziși și termenii medicali specializați este mai dificilă din perspectiva limbii române ca L2, necesitând criterii suplimentare, cum ar fi raportarea la CECRL[8] sau complexitatea lingvistică a termenilor (etimologie, derivare, opacitate semantică). Terminologia medicală generală ar trebui să includă termenii necesari unui vorbitor străin pentru situații medicale cotidiene, fără a intra prea mult în terminologia medicală de specialitate. Atribuirea pe nivel este relativ simplă aici: *medic*, *durere* și *febră* pot fi incluși la nivelul A2, *tratament* și *urgență* la nivelul B1, în timp ce *diagnostic* sau *transfuzie* țin de B2-C1. Un termen precum *atrioventricular* se încadrează simultan la termeni anatomici, când denumește conexiunea sau structura dintre atrii (în *sept atrioventricular*), dar și la termeni specializați, când se referă la patologii și afecțiuni. De aceea, a fost nevoie de o rafinare suplimentară a analizei, pe baza contextelor, păstrând doar denumirile de structuri anatomice la termeni anatomici și excluzând alte categorii. Denumirile care țin de biologia celulară, precum *colagen*, *keratină*, *mielină* etc., termenii care denumesc procese: *contractilitate*, *excitabilitate* etc. sau denumirile anatomotopografice: *distal*, *proximal* etc. nu au fost incluse. În același timp, am reținut în categoria termenilor specializați termenii descriptivi adjectivali, de tipul: *cerebrospinal*, *rahidian*, *plantar*, *scapular* etc., care presupun cunoașterea denumirilor de structuri anatomice cu care sunt relaționați. Termenii anatomici cu frecvențe mai mari sunt: *corp* (182), *sânge* (100), *organism* (98), *inimă* (82), *organ* (161), *sistem* (124) și *țesut* (81).

Orientarea spre teme medicale generale a condus, în corpusul Argăseală *et. al*., la prezența unor denumiri care țin de diagnosticare și farmacologie: *apendicită* (45), *paradontoză* (25), *rahitism* (21), *antibiotic* (50), *administrare* (40). Un număr mare de ocurențe au termenii anatomici *mușchi* (155), *os* (129), *piele* (112), *corp* (111), *sânge* (108), *inimă* (100), *organ* (92), *unghie* (89). În același timp, am putut constata absența denumirilor unor părți ale corpului: *pleoapă, geană, sprânceană, antebraț, nară* etc. sau o singură ocurență pentru *coapsă* și *călcâi*. În categoria terminologiei medicale generale, se remarcă: *medic* (113), *boală* (75), *medicament* (71), *pacient* (67), *durere* (49) și *tratament* (26). Pe eșantionul de 1000 de tokenuri, am obținut un inventar cu 132 de termeni, în cadrul căruia distribuția este: 43% termeni anatomici, 40% termeni specializați și 17% termeni de uz general.

[8] Cadrul european comun de referință pentru limbi.

Centrat pe aspecte de anatomie și fiziologie a corpului uman, cu funcții, aparate și sisteme, corpusul Dafinoiu conține 5042 de tokenuri, cu următoarea distribuție pentru cei 168 de termeni medicali rezultați din eșantion: 48% termeni anatomici, 18% termeni de uz general și 34% termeni specializați. Dintre termenii anatomici, cele mai multe ocurențe prezintă: *sânge* (104), *os* (96), *organ* (73), *celulă* (63), *inimă* (46), *plămân* (39), *stomac* (28), *venă* (26), *intestin* (23) și *arteră* (23). Un număr important de ocurențe se înregistrează la verbe legate de prezentarea structurilor anatomice: *a forma* (77), *a se afla* (67), *a se găsi* (51) etc. Cuvintele de uz general evidențiate la corpusurile anterioare înregistrează un număr mai mic de ocurențe sau nu apar deloc: *medic* (10), *doctor* (4), *boală* (2), *tratament* (2), *bolnav* (0), *medicament* (0).

Corpusul cumulat a condus la un inventar de 3515 tokenuri comune, termenii anatomici cu frecvențe agregate mari fiind: *os* (250), *mușchi* (220), *sânge* (200), *organ* (180), *piele* (110), *inimă* (110), iar dintre cei de uz medical general: *medic* (140), *boală* (90), *sănătate* (50) și *tratament* (70).

4. REZULTATE

Analiza întreprinsă asupra corpusului electronic SiMoNERo și asupra corpusurilor constituite pe baza unor cursuri academice de limbaj medical permite formularea unor observații relevante atât pentru descrierea vocabularului medical în limba română, cât și pentru selecția lexicală destinată predării limbii române ca L2. Corelarea datelor cantitative cu analiza calitativă a unităților lexicale evidențiază caracterul stratificat al vocabularului medical, în care vocabularul medical general, terminologia anatomică și terminologia specializată se află într-o relație de continuitate și interdependență, mai degrabă decât într-una de delimitare categorică.

Datele furnizate de SiMoNERo, ca reprezentare a discursului medical specializat, indică prezența consistentă a celor trei categorii urmărite: termenii medicali specializați dețin ponderea cea mai mare (41%), urmați de termenii anatomici (33%) și de termenii de uz medical general (26%). Această distribuție evidențiază coexistența, în același discurs, a unor unități lexicale cu grade diferite de specializare și confirmă faptul că discursul medical specializat nu se construiește exclusiv prin intermediul terminologiei propriu-zise. Lexeme precum *pacient, boală, tratament* sau *terapie* participă în mod esențial la construirea sensului și la asigurarea coerenței comunicării de specialitate.

În raport cu SiMoNERo, corpusurile academice relevă o deplasare a centrului de greutate către terminologia asociată conținutului disciplinar. În corpusurile Biriș și Dafinoiu, orientate predominant către anatomie și fiziologie, termenii anatomici constituie componenta lexicală dominantă, în timp ce corpusul Argăseală *et al*., caracterizat printr-o tematică medicală mai diversificată, prezintă o distribuție mai echilibrată între termenii anatomici și cei specializați. Aceste diferențe susțin

caracterul dependent de domeniu, tematică și gen discursiv al vocabularului medical și arată că inventarul lexical relevant nu poate fi stabilit adecvat în absența raportării la specificul materialelor abordate.

Un rezultat important al cercetării îl constituie dificultatea stabilirii unor granițe ferme între termenii anatomici și termenii specializați. Mobilitatea semantică a unor elemente lexicale și dependența statutului lor de contextul de utilizare susțin interpretarea vocabularului medical ca un continuum de specializare. Din perspectivă didactică, această observație este cu atât mai relevantă cu cât statutul terminologic al unei unități nu coincide în mod necesar cu gradul său de dificultate pentru un vorbitor de română ca L2. Termeni precum *inimă, os, mușchi, sânge* sau *plămân* pot constitui un nucleu anatomic relativ accesibil, în timp ce unități precum *miocard, tricuspid* sau *fetal* presupun un grad mai ridicat de specializare conceptuală și terminologică.

Analiza corpusurilor academice evidențiază, totodată, importanța lexicului verbal și a unităților lexico-sintactice recurente pentru comprehensiunea discursului medical. Verbe precum *a se afla, a se găsi, a prezenta, a cuprinde, a forma* sau *a alcătui*, deși nu au statut terminologic, contribuie la exprimarea relațiilor de poziție, structură și funcție și, prin frecvența lor, participă la configurarea discursului academic medical. Prin urmare, selecția lexicală pentru româna ca L2 nu ar trebui să se limiteze la inventarierea substantivelor specializate, ci să includă și unitățile lexicale și lexico-sintactice recurente care susțin înțelegerea și producerea discursului de specialitate.

Compararea corpusurilor a permis identificarea unui nucleu lexical comun, reprezentat de unități precum: *os, mușchi, sânge, organ, piele, inimă, medic, boală, sănătate* și *tratament*. Stabilitatea acestei componente lexicale nu exclude însă variația determinată de domeniul medical și de tipul discursului. Rezultatele arată, astfel, că frecvența de corpus nu poate constitui singurul criteriu pentru selecția lexicală. Relevanța didactică presupune corelarea frecvenței cu factori precum transparența semantică, gradul de specializare, utilitatea comunicativă, apartenența la un anumit domeniu medical, recurența contextuală și nivelul de competență lingvistică al cursanților.

Din această perspectivă, raportarea la nivelurile de competență prevăzute de CECRL poate reprezenta un criteriu suplimentar pentru rafinarea inventarului lexical. Datele analizate susțin existența unui nucleu lexical medical fundamental, situat la intersecția vocabularului general, a vocabularului anatomic și a terminologiei medicale specializate. Acest nucleu trebuie înțeles nu ca o categorie închisă și invariabilă, ci ca un repertoriu dinamic, a cărui componență variază în funcție de domeniul medical, de tipul discursului și de profilul cursanților. În acest cadru, SiMoNERo oferă imaginea vocabularului utilizat în discursul medical specializat, în timp ce corpusurile academice permit observarea modului în care acest vocabular este selectat, organizat și contextualizat în procesul de transmitere a cunoștințelor medical-lingvistice.

Cercetarea evidențiază, de asemenea, utilitatea metodologiei bazate pe corpus pentru fundamentarea selecției lexicale. Instrumente precum AntConc facilitează identificarea frecvențelor, a grupurilor lexicale și a contextelor de utilizare, oferind o bază empirică pentru deciziile de selecție. Analiza automată nu poate însă substitui interpretarea lingvistică și didactică: lematizarea și gruparea formelor, stabilirea statutului lexical și semantic al unităților, precum și evaluarea gradului de specializare presupun intervenția cercetătorului. Se conturează, prin urmare, necesitatea unei metodologii mixte, în care analiza cantitativă furnizează datele empirice, iar interpretarea calitativă le transformă în informație relevantă pentru proiectarea didactică.

În consecință, rezultatele cercetării susțin necesitatea conceperii vocabularului medical pentru româna ca L2 pe baza unei combinații de criterii: frecvență, relevanță, grad de specializare, accesibilitate și progresie lingvistică. O asemenea abordare presupune trecerea de la simpla selecție a termenilor cu frecvențele cele mai ridicate la identificarea relațiilor dintre unitățile lexicale, a contextelor recurente și a funcțiilor acestora în discursul medical și academic. Din această perspectivă, progresia didactică poate porni de la vocabularul medical general și de la termenii anatomici cu un grad mai ridicat de accesibilitate și poate avansa către terminologia specializată, în funcție de nivelul de competența lingvistică și de nevoile de formare ale cursanților.

5. CONCLUZII

Contribuția principală a studiului a constat nu doar în identificarea unor inventare lexicale, ci și în formularea unor principii pentru construirea unui inventar lexical medical destinat vorbitorilor de română ca L2. Datele de corpus oferă un punct de plecare empiric pentru identificarea vocabularului relevant, în timp ce analiza semantică, terminologică și didactică permite adaptarea acestuia la nivelul de competență și la nevoile academice ale cursanților. Selecția lexicală devine, în acest fel, un proces gradual și argumentat empiric, situat la intersecția dintre lingvistica de corpus, terminologie și didactica limbilor străine.

În ansamblu, rezultatele arată că predarea vocabularului medical nu poate fi redusă la transmiterea unei nomenclaturi de specialitate. Accesul la discursul medical academic presupune însușirea unui repertoriu lexical stratificat, în care vocabularul general, denumirile anatomice, termenii specializați și unitățile lexicale recurente ale discursului se articulează într-o rețea funcțională. Această perspectivă integratoare poate constitui baza pentru elaborarea unor materiale didactice și a unor inventare lexicale adaptate specificului românei ca L2, precum și pentru validarea ulterioară a acestora prin raportare la nivelurile de competență lingvistică și la nevoile concrete ale formării medicale în limba română.

BIBLIOGRAFIE

Argăseală G. *et. al*., 2022, *Limba română pentru străini: limbaje specializate, vol. I: Științe medicale*. Brașov: Editura Universității *Transilvania*.

Bălănescu, O., 1998, *Limbaj medical românesc pentru străini*. București: Ariadna'98.

Biriș, G., 2015a, „Vocabular medical pentru Anul pregătitor: inventar, dificultăți, perspective", în F. Sâmihăian, M. Spătaru-Pralea (coord.), *Limba și literatura română – Perspective didactice. Actele celui de-al 14 si al 15-lea Colocviu internațional al Departamentului de lingvistică (București, 29 noiembrie 2014 și 28 noiembrie 2015)*, București: Editura Universității din București, 141-151.

Biriș, G., 2015b, *Limbaj medical. Anatomie: curs pentru studenții străini din anul pregătitor*. București: Editura Universității din București.

Dafinoiu, C., 2023, *Anatomia și fiziologia omului. Manual de limbaj medical pentru studenții străini din anul pregătitor*. Constanța: Ovidius University Press.

Faur E., A-N. Pop, 2017: „Tradițional vs modern în lexicografie. O aplicație pe corpusul *Romanian Web 2016"*, *Caietele Sextil Pușcariu III*, Cluj-Napoca, 201–223, https://csp.inst-puscariu.ro/pdf/CSP%20III/CSP_III%20[Pages%20201%20-%20223].pdf.

Ferguson, G., 2001, „If you pop over there. A corpus-based study of conditionals in medical discourse", *English for Specific Purposes* 20, 61–82.

Ion, R. *et al.* 2012: „ROMBAC: The Romanian Balanced Annotated Corpus", în N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (ed.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. Istanbul: European Language Resources Association, 339–344, http://www.lrec-conf.org/proceedings/lrec2012/pdf/218_Paper.pdf.

Mișcalencu, D. *et al.* 1983, *Anatomia omului. Manual pentru studenții străini din anul pregătitor*. București: Editura Didactică și Pedagogică.

Mitrofan, M. și Tufiș, D., 2018: „BioRo: The Biomedical Corpus for the Romanian Language", în *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan: European Language Resources Association (ELRA), 1192–1196, https://aclanthology.org/anthology-files/pdf/L/L18/L18-1191.pdf.

Mîrzea-Vasile, C., 2017: „Corpusurile de limba română și importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină", *Romanian Studies Today* 1, 74–95.

Șoșa, E., I. Marinescu, E. Petre, 1990, „Organizarea vocabularului medico-farmaceutic în scop didactic", *Limba română* XXXIX, 5–6, 437–445.

Salager-Meyer, F., 1983, „The lexis of fundamental medical English: Classificatory framework and rhetorical function (a statistical approach)", *Reading in a Foreign Language*, 1(1), 34–64.

Salager-Meyer, F., 1987, „Infinitive clauses in medical English literature: A rhetorico-grammatical approach", *EMP Newsletter* 4/1, 23–33.

Salager-Meyer, F., 1992, „A text-type and move analysis study of verb tense and modality distribution in medical English abstracts", *English for Specific Purposes* 11, 93–113.

Velican-Aldea, M. *et al.*, 1991, *Limba română pentru studenții străini. Profil medical*. București: Editura Didactică și Pedagogică.

ANEXA 1

1. TA: arteră (102), auditiv (23), cord (54), ficat (35), muscular (34), oral (49), organism (19), osos (24), renal (152), suprarenal (26), venos (61), sistem (98), țesut (88), vas (51), membrană, (47), femural (33), abdominal (32), aparat (25), hormonal (21), nervos (19), adipos (19);

2. UG: afecțiune (90), boală (335), diagnostic (200), durere (41), examen (33), infecție (50), inflamație (23), medic (49), medicament (50), pacient (1034), semn (50), simptom (82), sânge (122), terapie (190), tratament (453), sănătos (30), urgență (27), bolnav (125);

3. TS: cardiac (199), vascular (122), cardiovascular (117), genă (102), ventricular (98), aortic (96), coronarian (93), leziune (92), pulmonar (91), tulburare (85), secreție (84), miocardic (78), metabolic (73), cerebral (61), hepatic (56), valvă (46), urinar (45), tiroidian (44), sinusal (40), atrioventricular (37), pancreatic (35), endotelial (31), circulație (29), proximal (26), fetal (26), sângerare (24), tisular (24), gastrointestinal, 22, tricuspidă (21), glomerular (21).

MEDICAL VOCABULARY FOR ROMANIAN AS A SECOND LANGUAGE (L2) FROM THE PERSPECTIVE OF ELECTRONIC CORPORA

Abstract

The study examines medical vocabulary in Romanian from the perspective of lexical selection for teaching Romanian as a second language (L2), based on a specialized medical electronic corpus and on corpora compiled from academic medical courses. The quantitative and qualitative analysis highlights the stratified nature of medical vocabulary, in which general medical vocabulary, anatomical terminology, and specialized terminology are interconnected along a continuum rather than being strictly delimited categories. The findings show that frequency alone is not sufficient as a criterion for lexical selection, as didactic relevance also requires consideration of the degree of specialization, accessibility, and communicative function of lexical units. The study thus supports a corpus-based approach to lexical selection for Romanian as an L2, complemented by linguistic, terminological, and didactic criteria.