ON THE NECESSITY OF A LEARNER CORPUS - ROMANIAN AS A FOREIGN LANGUAGE (RFL)

Antonela Arieşan and Lavinia-Iunia Vasiu

"Babeş-Bolyai" University of Cluj-Napoca

Abstract: Nowadays the field of corpus linguistics (CL) is continuously developing and its applicability has started to be recognized in various domains, due to CL's ability of clarifying, confirming or denying certain theories or just by offering examples of natural language opposite to the ones specially built to serve the purpose demonstrated. Among these fields a special type of corpus can be identified – learner corpus – (containing foreign speakers' written and/or spoken productions) which is recognized in the field of foreign languages as an essential resource in producing teaching, learning and assessment materials. Even though the utility of this type of corpus for Romanian as a foreign language (RFL) was many times recognized until this point a RFL learner corpus is not available yet. In this paper we try to approach this issue by presenting the most important characteristics of this special type of linguistic tool, the ways in which it is used in developing the materials needed in teaching and assessing different foreign languages and finally we will argue upon the necessity of a RFL learner corpus presenting its advantages in our particular case.

Keywords: Romanian as a foreign language, learner corpus, teaching/learning/assessment materials

0. Introducere

Remarcăm că, tot mai des, se vorbește despre utilitatea corpusului în demonstrarea diferitelor teorii dintr-o varietate de domenii de cercetare. Aria limbilor străine nu face excepție de la această tendință, întâlnind cu o frecvență sporită studii bazate pe corpus (oral sau scris), denumit în literatura străină de specialitate *learner corpus,* iar în limba română, *corpus al vorbitorilor non-nativi*.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

1. Corpusul lingvistic, în general

Corpusul lingvistic reprezintă obiectul de studiu al analizei lingvistice de corpus (*Corpus Linguistics*), pe care o vom denumi în continuare LC şi care a devenit cunoscută pe scară largă abia în anii 1980[1]. După Tony McEnery şi Andrew Hardie, domeniul LC este unul eterogen, care se axează asupra unui set de proceduri şi metode pentru studiul limbii. Chiar dacă unele dintre procedurile utilizate sunt extrem de clare, există altele care se află încă în stadiu de dezvoltare şi tocmai de aceea pot să ne ajute în rafinarea unor teorii sau în confirmarea altora care până acum nu se puteau realiza din cauza dimensiunii prea mari a materialului.[2] Oricum, prin utilizarea celor două mari tipuri de analiză, oferite lingviştilor de LC, analiza calitativă (*concordanţele* – cuvintele văzute în context) şi cea cantitativă (*listele de frecvenţă*), aceştia pot găsi răspunsuri la un set de întrebări prestabilite. Astfel, tot mai frecvent se recomandă utilizarea corpusului pentru verificarea şi oferirea unor exemple din limba naturală, în defavoarea acelora din textele literare. „Orice teorie lingvistică azi trebuie susţinută de o evaluare care să se ancoreze într-un corpus semnificativ de exemple. Corpusurile şi metodele de evaluare au devenit atât de semnificative în lingvistica zilelor noastre încât au început să fie organizate importante conferinţe internaţionale specific orientate pe această problematică."[3] Această observaţie este, desigur, valabilă şi în ceea ce priveşte domeniul RLS. O programă de curs, un manual, un dicţionar sau un test realizate pe baza unui corpus al vorbitorilor non-nativi sau validate cu ajutorul analizei acestuia, vor prezenta mai multă încredere decât un material bazat pe intuiţia de vorbitor nativ a autorului sau pe experienţa profesională a acestuia.

Cu privire la termenul *corpus*, acesta este definit în literatura de specialitate drept o colecţie de texte, caracterizată prin: dimensiune mare şi posibilitate de prelucrare electronică, realizată cu scopul de a fi reprezentativă pentru limba în cauză şi pentru a fi analizată din perspectivă lingvistică (calitativă şi cantitativă)[4]. Şi Tony McEnery şi Andrew Hardie afirmă că atunci când vorbim despre corpus, ne referim la un „un set de texte care de obicei sfidează analiza lingvistică cu mâna şi ochiul într-un timp rezonabil."

[1] *Michael McCarthy, Anne O'Keeffe, Historical perspective. What are corpora and how have they evolved in Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 5.*
[2] *Tony McEnery, Andrew Hardie, Corpus Linguistics: Method, Theory and Practice, Cambridge University Press, Cambridge, 2012, p. 1.*
[3] Dan Cristea, *Resurse lingvistice şi tehnologii ale limbajului natural. Cazul limbii române. Prelegerile Academiei Române. Filiala Iaşi, 2005*, p. 3.
[4] MichaelStubbs, *Language corpora*, in: Davies, Alan, Catherine Elder, *The Handbook of Applied Linguistics*, Blackwell Publishing Ltd, p. 106. şi O'Keefee, Anne, McCarthy, Michael, *From Corpus to Classroom*, Ronald Carter, Cambride University Press, Cambridge, 2007, p.1-2.

Arhipelag XXI Press, Tîrgu Mureş, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

(tr. n.)[5] Până a fi cunoscut așa cum a fost definit mai sus, „termenul *corpus* se referea la o colecție de texte similare. În secolul XVIII cercetătorii făceau referire *la un corpus al poeților latini* sau la *un corpus de legi*. Oricum, prima citare a acestui termen se pare că e datată în anul 1956, cu sensul de *corp de materiale scrise sau vorbite*, care stau la baza unei analize lingvistice." (tr. n.)[6] Utilitatea corpusului lingvistic și a analizei acestuia în cercetare nu a fost nicicând negată, însă perspectiva din care acestea sunt privite este diferită. Se disting, astfel, două categorii de cercetători. Pe de-o parte, unii consideră că LC este o *metodă* pe care ne putem baza studiile, care ne confirmă sau ne infirmă teorii, iar, pe de altă parte e cealaltă categorie, cea a cercetătorilor cunoscuți drept *neo-Firthians* (urmași ai lui J. R. Firth), care promovează LC ca pe o *teorie*, iar însuși corpusul lingvistic, în accepțiunea acestora, are statut teoretic.[7] Aceștia își bazează afirmația pe ideea conform căreia, folosind LC, obținem noi teorii lingvistice. De pildă, John Sinclair afirmă că „o astfel de abordare evită orice altă clasificare lingvistică anterioară, iar consecința ar fi că astfel de teorii sunt simple surse de distracție de la așa-numita *teorie ascunsă în spatele datelor*." (tr. n.)[8] Tony McEnery și Andrew Hardie oferă un argument puternic în defavoarea acestei teorii, aducându-l în discuție pe Noam Chomsky, care consideră că informațiile din corpus nu pot fi folosite drept o sursă de înțelegere a naturii limbajului. Astfel, dacă vedem interpretarea grupului *neo-Firthian* prin raportare la Chomsky, vom ajunge la concluzia că „nimic în afara informațiilor din corpus nu poate fi folosit ca sursă de înțelegere a naturii limbajului."[9] În accepțiunea noastră, LC este un mijloc de a ajunge la un rezultat, o metodă pe care o putem utiliza pentru a ne valida cercetările sau pentru a găsi răspunsuri la unele întrebări complexe, însă nu credem că aceasta poate constitui teoria în sine.

În funcție de tipul textului, modalitatea de transpunere, conținut, prelucrare sau dimensiuni, corpusurile pot fi de mai multe feluri. Vorbim, astfel, despre: corpusuri de dimensiuni mari sau de dimensiuni mici, corpusuri primare sau adnotate, corpusuri de referință sau specializate, corpusuri transcrise sau orale etc. Dintre acestea se distinge un tip special de corpus, este vorba despre corpusul vorbitorilor non-nativi (*learner corpus*).

[5] *Tony McEnery,Andrew Hardie, op. cit., p. 2.*
[6] *După Michael McCarthy, Anne O'Keeffe, op. cit., p. 5.*
[7] *Tony McEnery, Andrew Hardie, op. cit., p. 147.*
[8] *J. Sinclair, Trust the Text: Language, Corpus and Discourse, London, Routledge, 2004, p. 191.*
[9] *Tony McEnery, Andrew Hardie, op. cit., p. 148.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

2. Avantajele unui Corpusul al vorbitorilor non-nativi pentru RLS

Corpusul vorbitorilor non-nativi este definit drept „o colecție de producții scrise și orale ale vorbitorilor non-nativi." (tr. n.)[10] Acesta e utilizat, tot mai frecvent, în procesele de predare, învățare și evaluare ale unei limbi străine, atât în cadrul metodei cunoscute în literatura străină de specialitate sub denumirea de *data-driven learning* (*DDL*, în accepțiunea urmașilor lui Firth), cât și ca metodă empirică de verificare și confirmare a variatelor teorii. Modificări semnificative se pot observa în felul în care e percepută limba, datorită existenței și a utilizării unui corpus al vorbitorilor non-nativi: „se mută accentul de pe dominația vorbitorului nativ și se aduce *limba* vorbitorului non-nativ în centrul atenției, oferind posibilitatea, la nivelul clasei, să se folosească o *limbă* pe care o pot crea și pe care o pot folosi cei care o învață." (tr. n.)[11] Așadar, se pliază mult mai bine nevoilor grupului-țintă în discuție. După cum am putut remarca, corpusul vorbitorilor non-nativi este un pas înainte recent, dar foarte important în domeniul CL. Granger, un nume reprezentativ pentru acest domeniu, definește corpusul vorbitorilor non-nativi drept „o colecție de texte *autentice* produse de vorbitori străini sau de minorități. Aceasta arată că editurile, dar și universitățile din străinătate au început să colecteze și să analizeze producții ale vorbitorilor non-nativi pentru a produce materiale didactice bazate pe corpus." (tr. n.) [12] Se remarcă, așadar, atenția sporită acordată creării unor instrumente de predare și de evaluare eficiente și clar delimitate pe nivelurile din *Cadrul European comun de referință pentru limbi* (*CECR*), documentul cu cea mai mare importanță în procesele de predare-evaluare ale limbilor europene.

În ciuda faptului că pentru limba română ca limbă maternă există diferite corpusuri, electronice sau tipărite[13], care nu constituie subiectul comunicării noastre, în ceea ce privește domeniul limbii române ca limbă străină (RLS) nu putem afirma același lucru. Cu toate acestea, necesitatea și utilitatea unui astfel de material au fost remarcate: „Corpusurile lingvistice reprezentative sunt foarte importante în descrierea corectă a unei limbi. Reprezentând o mostră de limbă standard, neutră (varianta de limbă pe care o învață nenativii în mod ideal, abstract), utilizarea unui asemenea corpus este absolut necesară și

[10] *Michael McCarthy, Anne O'Keeffe, op.cit., p. 8.*
[11] *Ibidem, p. 8*
[12] Granger Apud Anne O'Keefee, Michael McCarthy, *From Corpus to Classroom*, Ronald Carter, Cambride University Press, Cambridge, 2007, p.24.
[13] Este vorba despre corpusuri ca: ROMBAC, RoCo_News, CoRoLa – The Reference Corpus of Contemporary Romanian Language etc. (în format electronic) și CORV, ROVA etc. (în format imprimat). Pentru o prezentare detaliată a acestora, a se vedea Carmen Mîrzea Vasile, *Corpusurile lingvistice de limba română și importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

în realizarea de materiale didactice pentru nenativi."[14] Dintre toate avantajele pe care un astfel de intrument le poate avea, se remarcă, așadar, acela de a ajuta la realizarea/de a sta la baza unor materiale didactice reprezentative, atât în ceea ce privește procesul de predare (manuale, auxiliare, dicționare, curricule, pe niveluri), cât și cel de evaluare (teste și grile de evaluare). Dacă româna ca limbă străină ar beneficia de un astfel de corpus, corelația dintre *ceea ce i se cere* studentului și *ceea ce poate realiza*, de fapt, va fi mult mai aproape de realitate. Cel mai elocvent exemplu în acest sens considerăm că este proiectul *English Profile*, bazat pe analiza *Cambridge English Corpus*, din care au rezultat *English Vocabulary Profile* și *English Grammar Profile*, două materiale extrem de utile în învățarea limbii engleze atât din perspectiva profesorului, cât și din cea a studentului. Aceste instrumente de cercetare oferă o inventariere clară pe nivelurile de referință din *CECR* a elementelor de vocabular utilizate de studenți la fiecare nivel, împreună cu exemple din producțiile acestora, ceea ce asigură o garanție în plus a echilibrului dintre obiectivele propuse în predare și cele din evaluare, material ce ar fi extrem de util și limbii române și irealizabil fără un corpus al vorbitorilor non-nativi[15]. Pe de altă parte, în *English Grammar Profile* structurile gramaticale sunt corelate, prin analiza corpusului, cu ceea ce e capabil studentul să utilizeze, fără a se apela, la experiență, la intuiție și la comparația cu alte limbi, ceea ce în cazul RLS încă are frecvent loc. Așadar, existând un astfel de material, s-ar putea asigura o uniformizare atât la nivelul unei singure instituții, cât și la nivel mai larg, la nivel național și european.Ce se întâmplă, de fapt, la această oră în cazul limbii române? Din păcate, din lipsa corpusului și a materialelor rezultate pe baza acestuia, fiecare cadru didactic, respectiv, centru de predare și evaluare, se raportează, în cel mai bun caz, la curriculumul din centrul său, sau la aspectele care sunt deja descrise[16], deși, de cele mai multe ori, intuiția pare a fi elementul hotărâtor. Pentru a ne susține afirmația, aducem în discuție o situație de dată recentă. În cadrul departamentului nostru a avut loc

[14] *Carmen Mîrzea Vasile, Corpusurile lingvistice de limba română și importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină, Institutul de Lingvistică al Academiei Române „Iorgu Iordan – Al. Rosetti" din Bucureşti /Facultatea de Litere, Departamentul de Lingvistică, Universitatea din București, 2014 (prezentare ppt)*

[15] *Pentru o argumentare asupra avantajelor unui inventar lexical al RLS, a se vedea articolul Observații asupra utilității inventarelor lexicale (Antonela Arieșan) în Ionuț Pomian (coord.), Înspre și dinspre Cluj. Contribuții linngvistice. Omagiu profesorului G.G. Neamțu la 70 de ani, Cluj-Napoca, Editura Scriptor și Editura Argonaut, 2015, p. 54-61.*

[16] *Ne referim aici la manualele deja existente, realizate în același mod, sau la:*
Moldovan, V., Pop, L., Uricaru, L., 2002, Nivel prag. Pentru învăţarea limbii române ca limbă străină, Strasbourg, Consiliul Europei.
Platon, E., Sonea, I., Vasiu, L., Vîlcu, D., 2014, Descrierea minimală a limbii române. A1, A2, B1, B2, Editura Casa Cărții de Știință, Cluj-Napoca.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

un atelier de validare a nivelului fiecărui exercițiu din banca de itemi, folosită la realizarea examenelor de RLS. În ceea ce privește exercițiile de vocabular și gramatică evaluate, din observațiile celor opt experți implicați (de diferite vârste și din diferite centre universitare), plasarea pe nivel s-a făcut, aproape exclusiv, în funcție de structurile gramaticale prezente în exerciții, deoarece pentru acestea au avut la dispoziție un instrument destul de clar și de detaliat pentru fiecare nivel[17], în vreme ce problemele de ordin lexical, de pildă, au constituit obiectul unor întrebări deschise, de tipul: „*a face plajă* e de nivel A1?" sau „*slujbă* mi se pare cam prețios arhaizant". Probabil că de aceea, chiar și în cazul exercițiilor de verificare a elementelor lexicale însușite, plasarea lor s-a făcut în funcție de structurile gramaticale sau de momentul la care structura în cauză este introdusă la instituția de proveniență (din țară sau străinătate): „am scris A2 la itemul doi, deoarece la cursurile noastre construcția «*a trebui să*» se învață în al doilea an de studiu, odată cu conjunctivul prezent", observații care într-o situație ideală în care experții ar beneficia de un cadru comun, bazat pe o analiză riguroasă a corpusului non-nativilor, nu ar trebui formulate, căci se recomandă ca nivelul să fie stabilit în funcție de niște descrieri precise pentru fiecare segment implicat (gramatică, lexic, acte de vorbire, tipuri de texte și funcții etc.).

În acest sens, aducem din nou în ajutor modelul limbii engleze, mai exact, principiul care a stat la baza realizării proiectului *English Profile Wordlists*: e mai important să se inventarieze limba pe care o știu și o folosesc vorbitorii din toată lumea decât cea vorbită de nativi. Ceea ce diferențiază EVP de alte studii de acest tip, așadar, este tocmai utilizarea diferitelor corpusuri pentru a susține cercetarea. „Descrierile precedente ale limbilor au fost realizate de specialiști care și-au folosit, în mare măsură, experiența de vorbitori și profesori. În contrast, deoarece folosește analiza corpusului, metodologia EVP este empirică, oferind dovezi clare în ceea ce privește cunoștințele de limbă pe care le au vorbitorii din diferite colțuri ale lumii, la fiecare nivel de competență lingvistică." (tr. n.)[18] În fine, avantajul utilizării unui astfel de corpus, conținând producții atât de diferite, realizate de persoane cu profile așa de diverse, este că ne poate oferi o

[17] Este vorba despre programa pe niveluri de la Departamentul de limbă, cultură și civilizație românească, iar, de curând, Elena Platon, Ioana Sonea, Lavinia Vasiu, Dina Vîlcu, *Descrierea minimală a limbii române. A1, A2, B1, B2*, Cluj-Napoca, Editura Casa Cărții de Știință, 2014. De remarcat că și acest material se bazează pe experiența profesorilor implicați, precum și pe comparația cu alte descrieri de acest fel consecrate altor limbi.

[18] *Harrison, Julia, Barker, Fiona (eds.), English Profile Studies. English Profile in Practice, Cambridge University Press, Cambridge, 2015, p. 4.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

descriere mult mai precisă a conținuturilor pe niveluri. „De asemenea, faptul că datele din corpus provin de la vorbitori din toată lumea, înseamnă că rezultatele obținute:

- nu sunt concentrate numai asupra limbii vorbite de nativi;
- nu sunt relevante numai pentru anumiți vorbitori, din anumite medii, întrucât limbile materne ale acestor vorbitori sunt atât de diverse cum sunt și vârstele și preocupările acestora." (tr. n.)[19]

Chiar și atunci când profesorul pregătește introducerea unor cuvinte, corpusul poate avea un rol însemnat, deoarece va ajuta la asigurarea unui control, dar și a unei progresii în predarea și, mai apoi, în testarea diferitelor sensuri ale unui element lexical. Astfel, se susține ideea conform căreia, folosind numai frecvența cuvintelor în textele scrise de vorbitorii nativi, nu vom ajunge întotdeauna la rezultate relevante. Tocmai de aceea, în urma calculului frecvenței, rezultatele trebuie trecute și printr-un corpus al producțiilor vorbitorilor non-nativi. „O analiză bazată numai pe frecvență, în cazul vorbitorilor nativi nu reține anumite cuvinte care le sunt folositoare cursanților și care au o frecvență mare în vocabularul folosit în clasă." (tr. n.)[20]. Utilitatea în cadrul orelor de RLS a unui corpus al vorbitorilor nativi, pe lângă cel al vorbitorilor non-nativi, nu e însă negată, întrucât profesorii care se ocupă cu predarea unei limbi nematerne sau străine ar putea fi interesați de mărimea vocabularului vorbitorilor nativi, deoarece această informație poate oferi unele indicații asupra sarcinilor pe care ar trebui să le propună pentru realizare, mai ales că vorbitorii de română ca limbă străină vor studia sau lucra alături de nativi.[21]

Indiferent de metoda aleasă, cercetătorii limbilor străine au concluzionat că „este evident nevoie ca elevul să cunoască primele aproximativ 3000 cele mai frecvente cuvinte din limbă. Acestea sunt considerate a fi prioritare și nu se recomandă învățarea altora înainte de a le cunoaște pe acestea."[22] Mai mult, Nation susține că, „după ce aceste 3000 de cuvinte au fost însușite, este mai important să ne concentrăm asupra dezvoltării unor strategii de înțelegere și învățare a cuvintelor cu frecvență mai redusă în limbă." (tr. n.)[23] Pe de altă parte, în susținerea acestei afirmații, vine și teoria lui Sinclair și Renouf, teorie legată de *lexical syllabus*, adică, capacitatea obținerii, din corpus, a unei liste de cuvinte, necesară profesorului în predare. Aceasta nu este, conform autorilor, o simplă listă de

[19] *Harrison, Julia, Barker, Fiona (eds.), op. cit., p. 4.*
[20] *Cambridge ESOL: Research notes, Issue 41, August 2010, Cambridge University Press, p. 3.*
[21] Paul Nation and Robert Waring, *Vocabulary size, text coverage and word lists*, p. 2.
[22] *Doar prin comparație, deocamdată, putem presupune și noi că această observație s-ar putea aplica și limbii noastre.*
[23] *Paul Nation and Robert Waring, op. cit., p. 4.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

cuvinte, ci o listă care reține diferitele aspecte ale limbii, bazată pe criteriul frecvenței. Aceștia susțin că „accentul ar trebui să se pună asupra a trei aspecte:

- cele mai cunoscute forme ale cuvintelor în limbă;
- tiparul de utilizare;
- combinațiile pe care acele cuvinte le formează." (tr. n.)[24]

Ideea din spatele acestei afirmații este că cele mai frecvente cuvinte au și o mulțime de sensuri. Așadar, nu este nevoie să se învețe foarte multe cuvinte, ci, mai ales, acele cuvinte care pot desemna multe alte idei. Un exemplu în acest sens poate fi verbul *a face*, cel mai prolific verb din limba română, care intră în combinații diferite, concrete (*a face o casă /curățenie/ mâncare/ tema*) sau mai abstracte (*a face descoperiri/ a face aranjamentele necesare*) etc.

Pe lângă criteriul frecvenței, amintit anterior și recunoscut în toate materialele ca necesar și important, Nation arată că deja în multe studii se ia în considerare și criteriul denumit *range* (al vecinătății), definită ca presupunând ocurența unui cuvânt în mai multe subsecțiuni ale unui corpus.[25] Aceasta nu înseamnă că trebuie să li se ofere elevilor liste interminabile de cuvinte, ci că autorii de manuale și de curricule, precum și profesorii ar trebui să dispună de astfel de liste pentru a decide dacă un anumit cuvânt merită atenție sau dacă un text avut în vedere este adecvat pentru a fi folosit în clasă sau nu.[26] Pe lângă toate aceste argumente privitoare la materialele de predare-evaluare, „un avantaj esențial pe care acest tip de corpus îl poate aduce în clasă este *autenticitatea*" (tr. n.)[27], deoarece se vor putea prelua exemple din corpus pentru a se ilustra contextele de utilizare ale cuvintelor sau ale structurilor gramaticale nou-introduse, exemple care, deocamdată, se oferă pe baza intuiției de vorbitor sau, la nivelurile inferioare (A1-A2) sunt construite artificial.

Corpusul în sine, din perspectiva DDL, poate avea utilitate imediată în cadrul orelor de RLS atât pentru profesori, cât și pentru studenți, în extragerea regulilor gramaticale sau de descoperire a diferitelor relații dintre anumite structuri avute în vedere în corpus. Gilquin și Granger arată că astfel „ei [studenții] devin mai implicați, mai activi

[24] *Tony McEnery, Course in Corpus Linguistics: Method, Analysis, Interpretation (curs on-line), Lancaster University.*
[25] I.S.P. Nation,*Learning Vocabulary in Another Language*, Cambridge University Press, Cambridge, 2001, p. 5.
[26] *Nation, op. cit., p. 9*
[27] *Gaëtanelle Gilquin and Sylviane Granger, How can data-driven learning be use in language teaching? in Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 359.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

și, în final, autonomi în procesul de învățare." (tr. n.) [28] În continuare, aceștia arată că utilizând astfel corpusul, studentul va câștiga mai multă încredere în sine, ceea ce nu poate decât să ajute în însușirea cât mai eficientă a limbii. În plus, considerăm noi, se obișnuiește să lucreze cu un material de mari dimensiuni, ceea ce în cazul studenților străini din anul pregătitor de limbă română poate constitui o adevărată provocare, mulți dintre ei atunci când intră la cursurile de română nu sunt obișnuiți să prelucreze texte, nici chiar în limbile lor. Având un corpus al vorbitorilor non-nativi profesorul, încă de la nivelul A1, ar putea lucra prin descoperire. De exemplu, identificarea diferențelor între genuri, la substantiv, identificarea formelor verbale, a adjectivelor sau a unor familii de cuvinte etc. Chiar și în scop autodidactic poate fi utilizat un astfel de material, de pildă, atunci când se caută exemple corecte cu unele cuvinte sau structuri nou-învățate și încă insuficient exersate. Gilquin și Granger evidențiază această funcție a corpusului (a celui adnotat, în acest caz) în special în ceea ce privește greșelile repetate făcute de vorbitor, în ciuda intervențiilor profesorului.[29] De fapt, recomandarea specialiștilor LC este să avem o variantă a corpusului „curată", fără adnotări, și una cu adnotări – ce se poate realiza automat - , deoarece nu toți cercetătorii folosesc în același fel și în același scop corpusul.

Tot în urma analizei lingvistice a unui corpus al vorbitorilor non-nativi, s-ar putea obține grile de evaluare valide, conținând chiar exemple clare preluate din corpus. La această oră, evaluarea, asemenea predării, se realizează extrem de diferit de la un centru universitar la altul. La Departamentul de limbă, cultură și civilizație românească de la Facultatea de Litere, Universitatea Babeș-Bolyai, se utilizează un set de grile pentru evaluarea producțiilor orale și scrise ale studenților, grile realizate de membrii departamentului, însă de nenumărate ori s-a enunțat necesitatea nuanțării acestora prin oferirea mai multor exemple de structuri pe fiecare nivel și pe fiecare bandă, de pildă.

Pe lângă avantajele pe care le poate aduce din perspectivă didactică corpusul vorbitorilor non-nativi, și din perspectiva cercetării RLS acesta este extrem de necesar, întrucât ar putea asigura, prin comparația cu un corpus al vorbitorilor nativi, o analiză mult mai profundă decât cele realizate până la această oră.[30] În plus, folosind un astfel de material se va putea valida ceea ce acum se crede/se recomandă a fi cunoscut la fiecare

[28] *Ibidem.*
[29] *Ibidem.*
[30] *Elena Tognini Bonelli, Theretical overview of the evolution of corpus linguistics in Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 25.*

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

nivel.[31] Astfel, o analiză detaliată a erorilor, dar și a structurilor stăpânite de vorbitorii non-nativi la fiecare nivel poate fi mai simplu efectuată. De aceea, este extrem de relevantă modalitatea în care materialul a fost colectat și introdus în format electronic. „Dacă textele reprezintă producții ale elevilor sau ale copiilor, e important să se păstreze redactarea non-standard și structurile gramaticale folosite inițial, chiar dacă sunt incorecte, deoarece vor fi extrem de relevante."(tr. n.)[32]

3. Concluzii

Materialele de predare, cele de evaluare, precum și cele de descriere, disponibile pentru româna ca limbă străină, chiar dacă sunt declarate pe nivelurile din CECR, sunt realizate pornind de la intuiția vorbitorului nativ, experiența de predare/evaluare în domeniul RLS și/sau de la comparația cu alte materiale existente pentru alte limbi, ceea ce conduce la instrumente didactice diferite, care nu pot garanta unitatea în predare și în evaluare la diferitele centre de limbi din țară și din străinătate. Se impune, deci, o bază comună – un corpus electronic al vorbitorilor non-nativi, care să poată asigura repartiția clară a elementelor lexicale, a structurilor gramaticale, dar și a variaților conectori sintactici pe fiecare nivel. Mai mult, așa cum am arătat în studiul nostru, un corpus al vorbitorilor non-nativi este indispensabil, în zilele noastre, pentru a asigura cercetări riguroase (comparative sau nu) în domeniul limbilor străine. Nu mai puțin util este corpusul în sine în cadrul orelor de curs, când poate constitui o sursă însemnată de exemple și texte pe niveluri, utilizabilă atât de profesor, cât și de student.

Neputând nega necesitatea unui corpus lingvistic reprezentativ pentru RLS am început colectarea producțiilor scrise (Antonela Arieșan), precum și a celor orale (Lavinia-Iunia Vasiu) în vederea realizării acestui corpus scris și oral al vorbitorilor non-nativi de RLS, folosit mai departe în obținerea unui inventar lexical al limbii române ca limbă străină pe niveluri, precum și a unor noi grile de evaluare a producțiilor orale RLS.

---

[31] Pentru o analiză a mijloacelor de asigurare a coeziunii și a coerenței discursului, bazată pe un corpus de producții orale ale studenților străini, prin comparație cu ceea ce li se predă/cere studenților la nivelul B2, dar și cu ceea ce utilizează nativii, a se vedea articolul Mijloace de realizare a coeziunii în RLS. Nivelul B2. Conectorii (Studiu de caz) (Vasiu, Arieșan) din prezentul volum.

[32] Randi Reppen, Building a corpus. What are the key considerations? în Michael McCarthy, Anne O'Keeffe (eds.), The Routledge Handbook of Corpus Linguistics, Routledge, New York, 2012, p. 35.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

Bibliografie:

1. ***Cadrul European Comun de Referinţă pentru limbi: predare, învăţare, evaluare*, (traducere din limba franceză de Gheorghe Moldovanu), Consiliul Europei, Diviziunea Politici Lingvistice, Strasbourg, 2003.

2. Capel, Annette, *Insights and issues arising from the English Profile Wordlists project* în *Research Notes*, Issue 41 August 2010, University of Cambridge, ESOL Examinations.

3. Cristea, Dan (2005), *Resurse lingvistice şi tehnologii ale limbajului natural. Cazul limbii române*. Prelegerile Academiei Române. Filiala Iaşi. (http://profs.info.uaic.ro/~dcristea/papers/cristea-prelegeri.pdf)

4. Harrison, Julia, Barker, Fiona (eds.), *English Profile Studies. English Profile in Practice*, Cambridge University Press, Cambridge, 2015.

5. McEnery, Tony, Hardie, Andrew, *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge, 2012.

6. Moldovan, Victoria, Pop, Liana, Uricaru, Lucia, *Nivel prag pentru învățarea limbii române ca limbă străină*, Consiliul de Cooperări Culturale, Strasbourg, 2001.

7. Nation, I.S.P., *Learning Vocabulary in Another Language*, Cambridge University Press, Cambridge, 2001.

8. Nation, Paul, Waring, Robert, *Vocabulary Size, Text Coverage and Word Lists*, 13 p (http://www.fltr.ucl.ac.be/fltr/germ/etan/bibs/vocab/cup.html)

9. O'Keefee, Anne, McCarthy, Michael, *From Corpus to Classroom*, Ronald Carter, Cambride University Press, Cambridge, 2007.

10. Platon, Elena, *Repere pentru delimitarea nivelurilor de competenţă lingvistică în RLS*, în G.G. Neamţu, Ştefan Gencărău, Adrian Chircu (eds.), *Limba Română – Abordări tradiţionale şi moderne*, Presa Universitară Clujeană, Cluj-Napoca, 2009, p. 499-505.

11. Platon, Elena, Sonea, Ioana, Vasiu, Lavinia, Vîlcu, Dina, *Descrierea minimală a limbii române. A1, A2, B1, B2*, Editura Casa Cărții de Știință, Cluj-Napoca, 2014.

12. Waring, Robert, Nation, Paul, *Second Language Reading and Incidentaly Vocabulary Learning*, 23 p. (http://www.robwaring.org/papers/various/waring_120304.pdf)

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)

Cursuri, prezentări ppt și pagini web:

1. McEnery, Tony, *Corpus Linguistics: Method, Analysis and Interpretation (curs on-line)*, Lancaster University, Marea Britanie.

2. *The English Profile Wordlists, Validation and future developments* (prezentare ppt): URL: http://www.academia.edu/1665365/The_English_Profile_Wordlists_validation_and_future_developments.

3. Mîrzea Vasile, Carmen, *Corpusurile lingvistice de limba română și importanța lor în realizarea de materiale didactice pentru limba română ca limbă străină*, Institutul de Lingvistică al Academiei Române „Iorgu Iordan – Al. Rosetti" din Bucureşti /Facultatea de Litere, Departamentul de Lingvistică, Universitatea din București, 2014 (prezentare ppt).

4. *English Vocabulary Profile*, URL: http://www.englishprofile.org/.

Arhipelag XXI Press, Tîrgu Mureș, ISBN: 978-606-8624-03-7

BDD-V4032 © 2016 Arhipelag XXI Press
Provided by Diacronia.ro for IP 216.73.216.77 (2026-11-18 10:55:29 UTC)