

Roxana ROGOBETE,  
Valentina MUREȘAN,  
Mădălina CHITEZ  
(Universitatea de Vest  
din Timișoara)

**Liste de cuvinte academice  
în limbile română și engleză:  
o analiză contrastivă  
bazată pe corpus**

**Abstract: (Academic Word Lists in English and Romanian: a corpus-based contrastive analysis)** For multiple reasons, there is a lack of studies on Academic Writing in Romanian language – few scholars deal with an overview of writing practices in Romanian HEIs, even fewer deal with the vocabulary or writing principles specific to certain disciplines. In this context, the Romanian “scientific production” is extremely varied at the discursive level, both because of the lack of standardised curriculum concerning Academic Writing and because of the heterogeneity of writing rules and publication conditions imposed by Romanian journals. The aim of this study is to analyse the relevance of *Academic Word Lists* for Romanian academic writing, bringing into discussion examples and methodologies from the English language sphere, where tools such as the one proposed by Avery Coxhead in the late 1990s have proven their effectiveness. The applied dimension of this paper is based on the analysis of an expert, bilingual corpus of scientific articles – *EXPRES* – compiled within the DACRE project (*Discipline-specific expert academic writing in Romanian and English: corpus-based contrastive analysis models*). The aim of the study is to propose and exemplify the possibility of constructing such a glossary based on writing practices, by examining some of the results provided by corpus linguistics tools such as N-Grams or concordance lines.

**Keywords:** *Academic Word Lists, EXPRES corpus, academic writing in Romanian, academic writing in English, corpus linguistics.*

**Rezumat:** Din multiple motive, spațiul românesc nu abundă în studii referitoare la scrierea academică în limba română – puține cercetări se ocupă de o panoramă a practicilor de scriere în mediul universitar, dar și mai puține se ocupă, diferențiat, de vocabularul sau principiile de redactare specifice unor discipline. În acest context, „producția științifică” românească este extrem de variată la nivel discursiv, atât din cauza faptului că nu există deprinderi unitare la nivelul curriculumului ce vizează practicile de scriere universitară, cât și din pricina eterogenității normelor de redactare și a condițiilor de publicare impuse de revistele românești de specialitate. Scopul acestui studiu este de a analiza relevanța unor liste de cuvinte academice (*Academic Word List*) pentru îmbunătățirea practicilor de scriere academică în română, aducând în discuție exemple și metodologii din sfera limbii engleze, în cazul căreia instrumente precum cel propus de Avery Coxhead la finalul anilor '90 și-au demonstrat eficiența. Dimensiunea aplicativă a lucrării de față se bazează pe analiza unui corpus expert, bilingv, de articole științifice – *EXPRES* – compilat în cadrul proiectului DACRE (*Scriere academică specifică disciplinei în limbile română și engleză: modele de analiză contrastivă bazate pe corpusuri lingvistice*). Miza studiului este de a propune și exemplifica posibilitatea construirii unui astfel de glosar pornind de la practica scriiturii, prin examinarea unor rezultate oferite de instrumentele lingvisticii de corpus, precum N-Grams sau linii de concordanță.

**Cuvinte-cheie:** *liste de cuvinte academice, corpusul EXPRES, scrierea academică în limba română, scrierea academică în limba engleză, lingvistică de corpus.*

## 1. Preambul: provocări... didactice

Studiul de față pornește, în fapt, de la dificultățile întâmpinate, din ce în ce mai frecvent, în mediul academic, în a oferi modele de bune practici în domeniul scrierii academice, modele destinate unor generații de „nativi digitali”. Pentru acești educabili, „migrarea de la pagină la ecran” (Relles, Tierney 2013, 478), provocată de emergența tehnologiilor utilizate atât în sfera educațională, cât mai ales în cea non-academică, privată, nu determină în mod automat eficientizarea și „profesionalizarea” practicilor de documentare ori de scriere. Mai ales în ultima perioadă, finalul anului universitar și sesiunile de finalizare a studiilor de licență și masterat aduc, inevitabil, noi discuții, cel puțin la nivelul comisiilor și al coordonatorilor, legate de practicile de scriere academică, cel puțin în domeniul științelor umaniste, în care astfel de sarcini de lucru reprezintă principala manieră de evaluare a performanței academice (vezi Tucan et al. 2020a). De la neglijențe în ceea ce privește formatarea și elementele de tehnoredactare, la aspecte care țin de elaborarea unui discurs coerent, academic și cu miză, discuțiile tind să scoată mereu în evidență minusurile pe care studenții le au în acest domeniu. Cât „investim” însă în îndrumarea lor – atât la nivel individual, cât și instituțional (local sau național)? Ce exemple de bune practici le putem pune la dispoziție, mai ales în cazul studenților care au toate tipurile de informații – valide și deloc valide – la un click distanță?

## 2. Scrierea academică în spațiul universitar românesc

Din multiple motive, spațiul românesc nu abundă în studii referitoare la scrierea academică în limba română – puține cercetări se ocupă de o panoramă a practicilor de scriere în mediul universitar (vezi Chitez et al. 2018), dar și mai puține se ocupă, diferențiat, de vocabularul sau principiile de redactare specifice unor discipline. Fiecare instituție de învățământ superior își redactează propriul regulament de finalizare a studiilor, fiecare facultate elaborează (intern sau în mod public) un ghid practic de redactare a licenței/disertației, specificând faptul că un candidat se supune cerințelor câte unui colectiv sau chiar coordonatorului – chestiune pe de o parte justificată, având în vedere că o lucrare filologică nu va fi similară uneia din zona IT, însă, pe de altă parte, dacă „îmbrățișăm” eticheta inter-/multi-/transdisciplinarității, am avea nevoie și de un *style sheet*, și mai ales de a-i face pe studenți să înțeleagă cerințele discursive ale stilului academic, științific. În fond, scopul scrisului academic ar fi tocmai *adaptarea la și integrarea în* comunitatea discursivă specifică disciplinei, urmărind, în același timp, o diseminare cât mai extinsă, pe plan național și internațional:

„Caracterul productiv al scrisului în mediul universitar ar însemna, cel puțin ca deziderat asumat, producerea de cunoaștere bazată pe cercetare, alături de claritate, corectitudine și flexibilitate în exprimare, ca deziderate retorice/stilistice. Dimensiunea sistematică ar reprezenta cunoașterea diferitelor tipuri de genuri și tipologii discursive implicate în scrierea academică (stadiul cercetării, metode utilizate, decupajul adecvat al materialului de analizat etc.) și utilizarea lor într-un

context disciplinar unitar (selectarea informațiilor și a cadrului conceptual dintr-un corpus relevant pentru domeniu).” (Tucan et al. 2020b, 74).

Însă putem observa inclusiv faptul că exemplele pe care noi, ca cercetători profesioniști, le putem oferi studenților sunt cât se poate de diverse – articolele din revistele de specialitate din diferite domenii nu vor fi nici ele unitare – nici din punct de vedere formal (vezi norme de tehnoredactare distincte de la o publicație la alta), nici din punct de vedere calitativ sau al structurii cerute pentru o astfel de lucrare științifică – precum există tipul IMRaD (*Introduction, Methods, Results, and Discussion*) folosit sistematic în engleză (cum, de altfel, despre distincții în cercetarea filologică, în spațiul românesc și est-european vs cel occidental, am arătat în Rogobete, Chitez 2019). Nu este deloc neglijabil faptul că există discrepanțe în acest sens inclusiv la nivelul filologiei românești. Pe de o parte, putem aminti publicații bine indexate, dar care funcționează ca „cerc închis”, sau care își pierd „ratingul” în scurt timp. Pe de altă parte, apar reviste la care nu se depune efort pentru indexări (de cele mai multe ori, din lipsa unor resurse consistente – umane și financiare), dar au sistem de peer review bine pus la punct. Există, în acest context, extrem de puține ghiduri specializate pe discipline – nu numai vizând normele, dar și vizând vocabularul specific sau terminologia de specialitate – chestiuni care țin, în fond, și de transparența în ceea ce privește criteriile de evaluare și „așteptările” comunității academice în cazul unor astfel de lucrări. Mai mult, putem aduce în discuție inclusiv decalajul dintre aceste așteptări ale cadrelor didactice și viziunea educabililor înșiși (vezi Băniceru, Tucan 2018).

### 3. Scrierea academică în limba engleză

Dacă în cazul limbii române discutăm în acest caz despre nevoia de a specializa și profesionaliza vorbitori nativi, în cazul limbii engleze resursele sunt multiple și elaborate în special pentru zona pedagogiei predării englezei ca limbă străină, nu maternă.

Cum demersurile pentru introducerea unor discipline separate de scriere academică au fost făcute mult mai devreme în spațiul anglofon, iar engleza e *lingua franca* predată la nivel internațional, instrumente în zona frazeologiei sau a vocabularului academic există, inclusiv pe bază de corpus. Astfel, există deja numeroase corpusuri academice, accesibile sau nu publicului larg, pe care le vom inventaria sumar:

- corpusuri orale, constituite din transcrieri ale unor prelegeri sau cursuri de la universități din UK sau SUA:
- BASE (British Academic Spoken English, <https://www.sketchengin.eu/british-academic-spoken-english-corpus/>) – accesibil publicului larg, acesta conține cursuri și seminare înregistrate la două universități (Warwick, Reading) în perioada 2000-2005, din domenii precum arte și științe umaniste, științe ale vieții, fizică, științe sociale;

- MICASE (Michigan Corpus of Academic Spoken English, <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>) – conține transcrieri ale discursurilor academice înregistrate la Universitatea din Michigan;
- corpusuri scrise, colectate în diferite moduri incluzând diferite discipline:
  - PICAIE (Pearson International Corpus of Academic English, <https://www.sketchengine.eu/picaie-corpus/>) – un corpus din 2010, disponibil cu acces limitat; în acest caz, majoritatea textelor sunt colectate din surse online, dar și din manuale de tip *Longman Higher Education*, ori din British National Corpus (BNC; inclusiv secțiuni academice ale BNC), pe domenii (conține patru domenii principale: științe umaniste, științe sociale, științe naturale și exacte, științe aplicate);
  - OCAE (Oxford Corpus of Academic English, <https://www.sketchengine.eu/oxford-corpus-of-academic-english/>) – acesta conține texte din reviste academice, dar și cursuri universitare din științe umaniste, științe ale vieții, fizică, științe sociale;
  - BAWE (British Academic Written English corpus, <https://www.sketchengine.eu/british-academic-written-english-corpus/>) – este un corpus open, ce conține lucrări academice scrise de educabili de la diferite niveluri de predare (ciclurile de licență și master), din cadrul universităților din Marea Britanie, pe domenii similare cu cele anterioare: științe umaniste, științele vieții, fizică, științe sociale;
  - CAEC (Cambridge Academic English, disponibil cu acces limitat și el, <https://www.sketchengine.eu/cambridge-academic-english-corpus/>) – este un corpus colectat de la universități din SUA și Marea Britanie, conținând cursuri, seminare, reviste, dar și prezentări studențești sau eseuri.

Vedem că ultimele trei corpusuri amintite nu fac diferența între „producătorii” în curs de specializare și cei experți, ceea ce poate fi o problemă, cel puțin din perspectiva celor care lucrează exclusiv cu metodologie de corpus. De altfel, în proiectele derulate în cadrul Facultății de Litere, Istorie și Teologie și CODHUS (*Centre for Corpus Related Digital Approaches to Humanities*), au fost elaborate două corpusuri distincte, unul studentesc, al textelor scrise de cei aflați în curs de formare, *Corpus of Romanian Academic Genres* (ROGER: <https://roger-corpus.org/>); altul de tip expert, al textelor scrise de cadre didactice sau cercetători cu afiliere instituțională, deja integrați în sistem (nivelul minim inclus aici este cel al cercetătorilor-doctoranzi care scriu în colaborare cu coordonatorul; *Corpus of Expert Writing in Romanian and English*, EXPRES: <https://expres-corpus.org/>).

Un corpus similar lui EXPRES și cel mai apropiat de scopul studiului nostru este Corpus of Academic Journal Articles (CAJA), elaborat de Iztok Kosem în 2010, care propune și un dicționar al englezei academice. Rămânând în această sferă a vocabularului academic, pentru limba engleză există cercetări similare din zona lexicografiei, a frazeologiei, care propun diferite liste de cuvinte sau structuri ce pot fi

inserate în discursul academic – aspecte care țin de retorică și coerența discursului mai ales:

- compilarea Academic Word List (Coxhead 2000), un inventar al termenilor ce apar în texte academice, dar nu în vocabularul general;

- Academic Collocation List (Ackermann, Chen 2013), care include colocații (îmbinări de cuvinte) prezente în textele academice, indiferent de disciplină;

- Academic Phrasebank (<https://www.phrasebank.manchester.ac.uk/>), o bază de date cu formulări potrivite pentru a organiza discursul academic (de la introducerea argumentelor/contraargumentelor, a exemplificărilor, la tranziții între diferite părți ale lucrării etc.)

#### 4. Cum rămâne cu „importul” metodelor?

Ne putem întreba pe ce se fundamentează această paralelă realizată între scrierea academică în limba română, respectiv în limba engleză, de vreme ce spațiul autohton a fost mai degrabă tributar altor modele culturale. Justificarea rezidă în „primatul” englezei în comunitatea academică, aspect care se repercutează în ultimii 20-30 de ani și asupra cercetării din România. Astfel, inclusiv în cazul cercetătorilor români, există domenii în care producția științifică este de departe mai consistentă în limba engleză, în detrimentul românei. Accesul la publicul internațional devine mai important decât diseminarea utilizând idiomurile naționale. Într-adevăr, însă, în cazul unui import metodologic (cum dezbatem în studiul de față, pentru dezvoltarea vocabularului academic), putem preconiza cel puțin două probleme majore. Pe de o parte, ne confruntăm cu o lipsă de unitate la nivel terminologic în spațiul românesc pentru un eventual dicționar frazeologic, al „structurilor stabile” (Țeghiu 2004) sau al colocațiilor. În studiile de specialitate, circulă structuri sau denumiri precum „unitate frazeologică, frazeologism, locuțiune, locuțiune frazeologică, grup sintactic stabil, izolare, idiotism, discurs repetat, formula fixă, expresie, perifrază, expresie idiomătică, îmbinare stabilă/ constantă de cuvinte, grup sintactic stabil, unitate sintagmatică, formulă/ combinație fixă, sintagmă” (Cernea 2009, 12). Colocația, de exemplu, apare destul de târziu în lingvistica românească. Sunt cunoscute delimitările pe care le face profesorul Sârbu în 1987: colocația este „relația stabilită între constituenții unei sintagme pe baza intercondiționării lor semantice (lexical-frazeologice)” (fiind comparată cu coligația care se bazează pe intercondiționare „morfosintactică, structural-formală”, Sârbu 1987, 53). Aprofundări vin însă târziu, în anii 2000, de pildă, prin Liviu Groza, care menționează că structura este alcătuită dintr-o „bază” și un „colocativ [...] subordonat conceptului denumit de bază” (2005, 20). Totodată, prea puțin discutată este perspectiva cantitativă a ocurențelor unor unități multicuvânt, precum N-gramele, care sunt „elemente constitutive importante în discurs” (Biber, Barbieri 2007, 268-270). Pe de altă parte, un alt risc major este ca studenții să insereze eventuale rezultate ale unor instrumente digitale doar la modul automat și nefiltrat în propriul discurs.

## 5. Corpusul EXPRES și proiectul DACRE

Tocmai din nevoia de a identifica modul autentic în care sunt realizate cercetările în spațiul universitar românesc a fost inițiat proiectul DACRE (*Scriere academică specifică disciplinei în limbile română și engleză: modele de analiză contrastivă bazate pe corpusuri lingvistice*, proiect de cercetare exploratorie) derulat în perioada 2021-2022. Acesta urmărește să compenseze penuria cercetării lingvistice aplicate, adaptate contextului autohton, care folosește metode de corpus pentru a oferi un sprijin pentru scrierea academică specifică disciplinei. Cercetarea este bazată pe datele colectate în corpusul EXPRES (*Corpus of Expert Writing in Romanian and English*) compilate în cadrul proiectului. EXPRES conține, la rândul său, câteva sub-corpusuri, ce reprezintă douăsprezece seturi de date de scriere academică de tip expert, i.e. articole de cercetare, în limbile română (vorbitori nativi, cercetători afiliați unor instituții din România), engleză L1 (vorbitori nativi – afiliați unor cercetători din spațiul anglofon) și engleză L2 (nativi români afiliați universităților românești, dar care publică în limba engleză), din patru discipline: Lingvistică, Științe Politice, Științe Economice, IT. Fiecare subcorpus este alcătuit din 200 de texte colectate din reviste de specialitate, articole publicate majoritar în perioada 2017-2022.

Obiectivul principal al DACRE este constituit de crearea de modele metodologice de analiză contrastivă (română vs engleză), ce rezultă din investigarea corpusurilor construite: caracteristici ale scrierii academice specifice disciplinei (cum ar fi preferințele lexicale) și particularitățile de scriere academică românească (de pildă, unități lingvistice).

Care sunt etapele compilării unui astfel de corpus? Vom schița câteva aspecte legate în special de corpusul din limba română. Într-o primă etapă, a fost nevoie de selecția revistelor din cele patru domenii necesare. Criteriul selectat a fost prezența revistelor în baze de date precum Clarivate, EBSCO, ERIH – pe cât posibil, însă, pentru a colecta 200 de texte recente, ne-am orientat și către publicații din baze de date mai puțin vizibile sau „punctate”. Dificultățile cele mai mari au fost în domeniul IT, pentru că nu se scrie recent în limba română în această sferă (mai multe informații despre această selecție sunt menționate de Rogobete et al. 2021). Ulterior, a fost evident nevoie de descărcarea articolelor potrivite, proces care s-a făcut parțial automat prin web scrapping, parțial manual (mai ales pentru publicațiile cu acces pe bază de abonament instituțional). Textele au fost triate și apoi codate după un ghid elaborat în interiorul proiectului (bazat pe normele TEI simplificate, Figura 1), ce a presupus, în mare parte, anonimizare, curățare și codare a unor elemente precum citatele, formulele sau alte aspecte care nu ne interesau din punct de vedere discursiv.

Codebook (excerpt):

Code	Code description
{JOURNAL_TITLE}	Journal name (confidential)
{JOURNAL_DETAILS}	Journal details, e.g. DOI, editing house website, etc. (confidential)
{AUTHOR_DETAILS}	Author details, e.g. name, e-mail address (confidential)
{AUTHOR_BIO}	Author biography (confidential)
{ARTICLE_HISTORY}	Article processing stages (confidential)
{PHON}	Phonetics elements
{ETM}	Etymology elements
{GLS}	Glossary
{SIGNATURE}	Signature (confidential)
{ACK}	Acknowledgement
{PHON}	"Declaration of conflicting interests"
{REF_LIST}	Reference list
{REF}; {(REF), {REF},{REF}}	References
<APP>	Appendix
<COD>	Programming code
<CONT>	Table of contents
<DATE>	Date (confidential)
<FAC>	Faculty (confidential)
<FIG>	Diagram; Graph; Statistics; Conceptual Map
<FILEF>	File format
<FM>	Mathematical formula
<IMG>	Picture; Image
<LOCATION>	Location (confidential)

Figura 1 Informații despre adnotarea corpusului EXPRES.

Pentru diferitele analize am utilizat apoi atât instrumente de corpus (Sketch Engine, în special), cât și rularea unor programe de tipul Python. La finalul proiectului, corpusul va fi disponibil pe platforma EXPRES, pentru analize ce pot fi realizate de publicul larg (Figura 2).



Figura 2: Pagina principală a platformei EXPRES, [expres-corpus.org](http://expres-corpus.org)

Pe această platformă, utilizatorii vor avea acces la metadate referitoare la tipologia textelor, iar administratorii pot gestiona textele stocate în baza de date (Figura 3).

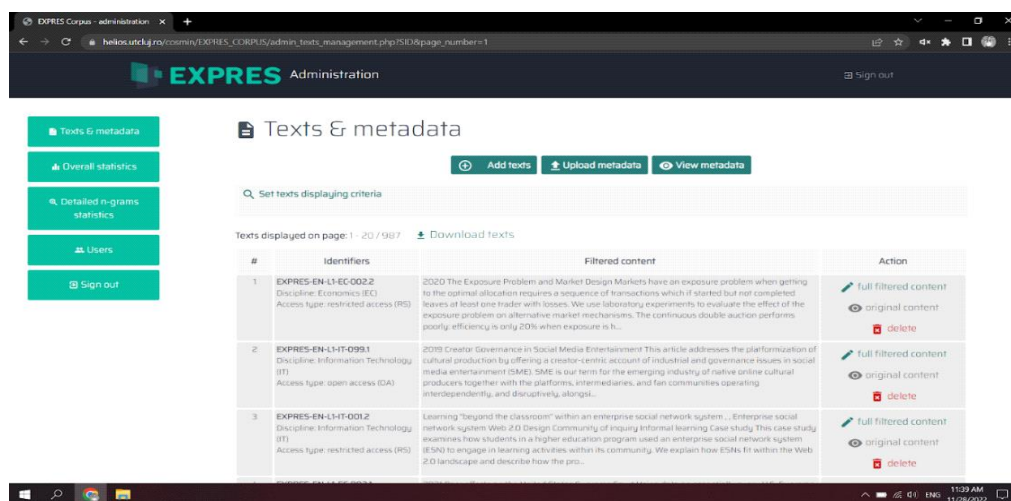


Figura 3: Pagina de administrare a metadatelor pentru corpusul EXPRES

Interfața frontend oferă utilizatorilor înregistrați posibilitatea de a căuta anumite cuvinte-cheie și de a rafina rezultatele obținute prin aplicarea unor filtre. Caracteristicile actuale ale platformei includ căutarea de termeni și sintagme, distribuția N-gramelor în corpus și vizualizări statistice pentru interogările efectuate. După introducerea unui termen/unei structuri în fereastra de căutare, utilizatorul poate filtra textele disponibile după: (i) limbă (engleză L1, engleză L2, română); (ii) domeniul (cele 4 domenii: Lingvistică, Științe Politice, Științe Economice, IT) (Figura 4).



The screenshot displays the search interface of the EXPRES corpus. On the left, there is a 'Search filters' panel with two sections: 'Language' and 'Discipline'. Under 'Language', 'Romanian - L1' is selected with a blue checkmark, while 'English - L1' and 'English - L2' are unselected. Under 'Discipline', 'Linguistics' is selected with a blue checkmark, while 'Economics', 'Information Technology', and 'Political Sciences' are unselected. At the bottom of the filters are 'Apply' and 'Reset' buttons. The main area is titled 'Search results' and shows 'Texts displayed on page: 1 - 20 / 39'. A list of six search results is shown, each with a numbered entry, the language and discipline, and a snippet of text with the word 'cercetare' highlighted in yellow. Each result has a '[show more]' link.

Figura 4: Interfața de căutare în corpusul EXPRES

## 6. Compilarea unei liste de cuvinte academice în limba română

Cum poate contribui proiectul DACRE în selectarea unei terminologii academice în limba română? Ce căutăm, așadar, în aceste articole de specialitate, la nivelul lexicului? Care sunt caracteristicile vocabularului academic în cele două limbi analizate – româna și engleza?

Pentru comunitatea de limbă engleză, competența în sfera limbajului academic presupune „cunoașterea și abilitatea de a folosi vocabular general și specific, structuri gramaticale specializate și complexe, structuri discursive și funcții lingvistice variate – toate cu scopul de a dobândi cunoștințe noi și de a dezvolta competențe, de a interacționa pe o anumită temă și de a împărtăși informații altora” (Bailey 2007, 10-11). În plus, Snow și Ucelli (2009, 119) au identificat cinci componente principale ale limbajului academic: poziționarea impersonală, încărcătura informațională, organizarea informației, alegerile lexicale și incongruența reprezentatională (trad. noastră). Astfel, în primă instanță, este nevoie de o poziționare „autoritară, detașată”, în care prevalează „expertiza în ceea ce privește conținutul” (Townsend et. al. 2009, 499). În schimb, în articolele din limba română asumarea personală (chiar și cu pluralul modestiei) este mult mai frecventă. Remarcăm diferențe și în privința „încărcăturii informaționale”. Pentru limba engleză, studiile menționează că este nevoie de „densitate și concizie”, căci „textele academice abordează de obicei concepte abstracte și/sau tehnice, se așteaptă ca autorii textelor academice sau vorbitorii să reducă numărul de cuvinte pentru exprimarea acestor concepte cât se poate de concis.” (Townsend et. al. 2009, 499). Nu credem că același lucru se poate spune și despre articolele din limba

română, pentru care revistele alocă, în zona lingvisticii, inclusiv peste 40 de pagini per articol. În ceea ce privește structurarea informației, am menționat deja nevoia unei argumentări, iar ordonarea logică a ideilor este, totuși, comună celor două sfere academice, deși putem constata că introducerea teoretică rareori respectă modelul CARS (vezi Chitez et al. 2022). Interesantă este chestiunea „alegerilor lexicale”, ce în engleză trebuie să reflecte „o mare diversitate lexicală și includ expresii formale și prestigioase” care presupun și „folosirea termenilor abstracți și tehnici, relevanți temei” (Townsend et. al. 2009, 499). Ultimul aspect, al „congruenței reprezentationale”, nu este deloc exploatat în limba română. Pentru cercetarea din limba engleză, acesta reprezintă „corespondența dintre limbaj și realitatea pe care o reprezintă” (Snow, Ucelli 2009, 121), iar studiile menționează câteva exemple de incongruențe, precum nominalizările, „care rezultă într-o densitate lexicală și informațională mare” sau într-o „folosire ca agenți a unor concepte abstracte, mai degrabă decât <actori> umani” (Townsend et. al. 2009, 499). Deși literatura de specialitate definește vocabularul academic din perspectiva predării, acesta ar trebui văzut ca un continuum, un proces, aplicabilitatea închizând acest cerc al remodelărilor lexicale. În ceea ce privește româna, pe lângă aspecte legate de concizie, claritate (v. <https://writing.unitbv.ro/stilul-academic/>), structurare a discursului, termenii specializați sau neologismele din diferite discipline devin importante. Așadar, vorbim mai puțin despre nivelul discursiv și mai mult despre un lexic specific. Însă eleganța exprimării în română nu este prescriptivă (a se vedea, în acest sens, și diversitatea tipologiei de articole prezente în reviste).

Să observăm, în fine, câteva diferențe practice de ordin tehnic, survolând instrumentele cele mai importante din cele două limbi analizate. Un Academic Word List precum cel elaborat de Averil Coxhead conține cele mai frecvente 570 de familii de cuvinte, nediferențiate din punct de vedere disciplinar, ci selectate în funcție de frecvența cu care apar într-un corpus de texte academice din arte, comerț, drept și științe. În schimb, *Micul dicționar academic* (2010) conține aproximativ 175 000 de termeni, sinteză a *Dicționarului limbii române*, dar la care se adaugă elemente neologice. Cum izvoarele utilizate sunt reprezentate și de texte vechi, sau texte de beletristică, cu siguranță o astfel de listă nu poate fi superpozată unui AWL. Ce este util însă în acest dicționar este menționarea domeniului în care este utilizat termenul respectiv. MDA este mai degrabă descriptiv, panoramic, oferă într-adevăr informații despre prima atestare, dar nu și despre frecvența de utilizare a unui termen, nu se limitează la vocabularul științific, la sensurile de bază, ci riscă să realizeze un amestec cu zona expresivității limbajului, cu sensurile figurate (chiar dacă exemplele sunt eliminate din dicționar). Ce justifică, așadar, caracterul științific, integral, al lexicului prezentat? Cum putem coagula niște liste de cuvinte academice pentru limba română?

O primă versiune ar fi să separăm pe domenii o eventuală ediție digitală a MDA, pe care ulterior să o „verificăm” pe un corpus de articole de specialitate – să confruntăm, practic, vocabularul „posibil” cu cel aflat în uz. Putem porni însă tocmai de la uz, de la compilarea acestui corpus EXPRES de articole științifice în limba română, un corpus în care să urmărim tocmai ocurențele unor cuvinte sau structuri mai

mult sau mai puțin stabile – structuri retorice, dar și cuvinte care țin de lexicul specializat. O astfel de analiză își găsește apoi aplicabilitatea pedagogică, pentru că rezultatele avute pe un corpus expert pot fi comparate cu analizele unor texte studențești, ori aceste liste de cuvinte academice pot fi integrate în predarea scrierii academice sau a cursurilor axate pe redactarea lucrărilor de licență/disertație. Rămân însă o serie de limitări ale studiului – când punem o astfel de „oglină” în fața practicilor de scriere academică de la nivel „înalt”, cât de „academică” este de fapt imaginea? În sensul în care, selectând lucrări în limba română, care nu vor fi în mod necesar cotate exclusiv în baze de date precum Web of Science, oare vor acoperi inclusiv domenii în care româna nu este „frecventabilă”, precum IT-ul – am reuși în fine să coagulăm un corpus relevant pentru scrierea academică specializată? Un alt aspect observat în compilarea corpusului a fost că ne-a fost dificil să găsim 200 de texte cu autori distincți. Strategic, EXPRES conține maximum 2 articole per autor unic (și nu e de neglijat tendința unor reviste – mai ales din zona Științelor Politice – de a publica aceiași specialiști în majoritatea numerelor). Misiunea DACRE este, prin urmare, nu tocmai ușoară.

Pentru a da o dimensiune aplicată analizelor noastre, am selectat cele mai frecvente cuvinte prezente în EXPRES în funcție de domeniu. Ne-am limitat la primele 7 substantive și adjective – care par a fi parțial specifice domeniului, respectiv la cele mai frecvente 10 alăturări de tip verb+prepoziție, structuri utilizate în special pentru a introduce exemplificări, observații finale, comentarii etc. (Figura 5).

substantive	expres_filogie	expres_economice	expres_politice	expres_it	EXPRES Economics	EXPRES Linguistics
	limbă	an	România	date	cuprinsă între (0.6471)	oprim asupra (0.6342)
	formă	risc	stat	sistem	clasificate drept (0.6278)	formată dintr- (0.6248)
	cuvânt	audit	an	utilizator	bazată pe (0.5644)	vorbi despre (0.5809)
	text	piață	parte	aplicație	bazează pe (0.5385)	vorbim despre (0.5593)
	sens	nivel	putere	informație	bazate pe (0.5332)	format dintr- (0.5504)
	termen	România	alegere	tip	adequate privind (0.5178)	caracterizează printr- (0.5416)
	substantiv	creștere	partid	model	bazat pe (0.5127)	bazată pe (0.5275)
adjective	expres_filogie	expres_economice	expres_politice	expres_it	EXPRES Political Science	EXPRES IT
	român	financiar	politic	mare	îndreptate împotriva (0.6423)	bazate pe (0.6242)
	vechi	economic	mare	nou	durat până (0.594)	bazat pe (0.6101)
	lingvistic	european	român	diferit	îndrepta spre (0.5654)	bazată pe (0.5959)
	românesc	mare	nou	utilizat	orientat spre (0.5555)	bazează pe (0.5423)
	semantic	nou	european	folosit	vorbi despre (0.5491)	referă la (0.5138)
	mare	național	social	bazat	discută despre (0.5286)	conduce la (0.4738)
	sintactic	principal	electoral	important	îndreptat împotriva (0.5277)	centrate pe (0.4712)
				vorbi despre (0.5274)	compusă din (0.466)	
				bazate pe (0.5006)	provenită din (0.466)	
				discuta despre (0.4914)	entități cu (0.4583)	

Figura 5: Cele mai frecvente substantive, adjective și structuri verb+prepoziție în EXPRES

Rezultatele precizate sunt preliminare, căci în cadrul DACRE urmează o etapă de asamblare a unui inventar – de combinații de cuvinte, colocații, termeni, aspecte pe care le putem păstra într-un glosar. Utilitatea unor astfel de demersuri se dovedește valabilă pentru, de pildă, depistarea unor eventuale combinații nepotrivite verb+prepoziție, care se observă mai ales la exercițiile de traduceri din limba engleză în limba maternă, când studenții nu au destulă experiență sau expunere la texte în română. Totodată, aceste erori sunt vizibile și în producerea de texte în limba română, când unele unități lexicale sunt preluate literal din engleză cu tot cu structurile combinatorii specifice, fără a verifica în limba română pentru a stabili uzanța, frecvența sau chiar acuratețea acestora (vezi cazul lui „a focusa” – DOOM 3, sau sensul mai recent al verbului „a centra” – tranzitiv, figurat – vezi Figura 5).

## 7. Concluzii și... alte provocări

Așadar, pentru a încerca adaptarea la modelele existente în spațiul academic de limbă engleză, cercetarea românească trebuie întâi să își găsească o coerență la nivel terminologic. Recunoaștem că nevoia de internaționalizare, de vizibilitate în cercetare este, de fapt, în detrimentul românei, considerată în fond „low-resourced language”. Astfel de instrumente precum cele oferite de platforma EXPRES și dezvoltate în cadrul DACRE – chiar și mai elaborate decât exemplele noastre – își pot găsi aplicabilitatea în cadrul unor cursuri de scriere academică, de formare de competențe de traducere, sau de lingvistică (pentru exemplificări sau exerciții de analiză), însă este nevoie de o foarte bună selecție a textelor din corpusul de referință. Trebuie să reținem și că orice paletă de instrumente de verificare pentru viitorii cercetători (glosare de termeni, dicționare de colocații, baze de date accesibile pentru căutări rapide) nu compensează lipsa de lectură, dar îi ajută pe studenți pentru eventuale nesiguranțe pe care le au în practicile de scriere academică. Cu toate aceste variabile, demersul pedagogic nu trebuie să exploateze instrumentele în mod izolat, ci cere o viziune holistică pentru a îmbunătăți experiențele de învățare ale studenților care vor fi viitorii cercetători.

## Referințe bibliografice

- Ackermann, K., Chen, Y.-H. 2013. *Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach*, in “Journal of English for Academic Purposes”, Volume 12, Issue 4, p. 235-247, <https://doi.org/10.1016/j.jeap.2013.08.002>.
- Băniceru, C., Tucan, D. 2018. *Perceptions about “Good Writing” and “Writing Competences” in Romanian academic writing practices: A questionnaire study*, in vol. M. Chitez, C.I. Doroholschi, O. Kruse, L. Salski, D. Tucan (Eds.), *University Writing in Central and Eastern Europe: Tradition, Transition, and Innovation*. Cham: Springer, p. 103-112.
- Biber, D., Barbieri, F. 2007. *Lexical bundles in university spoken and written registers*, in “English for Specific Purposes”, vol. 26, p. 263–286.
- Cerneș, M. 2009. *Unități frazeologice în limba română*. Târgu Mureș: Editura Kreativ.
- Chitez, M., Doroholschi, C.I., Kruse, O., Salski, L., Tucan, D. (Eds.). 2018. *University Writing in Central and Eastern Europe: Tradition, Transition, and Innovation*. Cham: Springer.

- Chitez, M., Mureșan, V., Rogobete, R. 2022. *How to Write Good Academic Papers: Using the EXPRES Corpus to Extract Expert Writing Linguistic Patterns*, in "Conference Proceedings. 12th International Conference the Future of Education", Hybrid Edition. Florence: Filodiritto Editore, p. 410-416, disponibil online la <https://conference.pixel-online.net/files/foe/ed0012/FP/3177-LANG5588-FP-FOE12.pdf>, ultima accesare: 10.X.2022.
- Coxhead, Averil. 2000. *A New Academic Word List*, in "TESOL Quarterly", vol. 34, no. 2, p. 213-238, <https://doi.org/10.2307/3587951>.
- Groza, L. 2005. *Dinamica unităților frazeologice în limba română contemporană*. București: Editura Universității din București.
- Kosem, Iztok. 2010. *Designing a model for a corpus-driven dictionary of academic English*. PHD thesis, Aston University, disponibil online la [https://publications.aston.ac.uk/id/eprint/14664/1/Kosem2010\\_484017\\_3.pdf](https://publications.aston.ac.uk/id/eprint/14664/1/Kosem2010_484017_3.pdf), ultima accesare: 10.X.2022.
- \*\*\* 2010. *Mic dicționar academic*. Vol. I: A-Me. Cuv. înainte de Eugen Simion, pref. de Marius Sala. București: Editura Univers Enciclopedic Gold.
- \*\*\* 2010. *Mic dicționar academic*. Vol. II: Mi-Z. București: Editura Univers Enciclopedic.
- Mureșan, M., Rogobete, R., Bucur, A.-M., Chitez, M., Dincă, A. 2022. *Phraseology in Romanian Academic Writing: Corpus Based Explorations into Field-Specific Multiword Units*, in A. Dinu, M. Chitez, L. Dinu, M. Dobre (Eds.), *Recent Advances in Digital Humanities. Romance Language Applications*. Berlin: Peter Lang, p. 29-48.
- Relles, S.R., Tierney, W.G. 2013. *Understanding the Writing Habits of Tomorrow's Students: Technology and College Readiness*, in "The Journal of Higher Education", 84:4, p. 477-505, DOI: 10.1080/00221546.2013.11777299.
- Rogobete, R., Chitez, M. 2019. *Disparities in Humanities Research Initiatives Within the European Higher Education Area*, in „Analele Universității de Vest din Timișoara. Seria Științe Filologice”, vol. LVII, Editura Universității de Vest din Timișoara, p. 145-153.
- Rogobete, R., Chitez, M., Mureșan, V., Damian, B., Duciuc, A., Gherasim, C., Bucur, A.-M. 2021. *Challenges in compiling expert corpora for academic writing support*, in "Conference Proceedings. 11th International Conference the Future of Education", Virtual Edition. Florence: Filodiritto Editore, p. 409-414, disponibil online la <https://conference.pixel-online.net/FOE/files/foe/ed0012/FP/3177-LANG5588-FP-FOE12.pdf>, ultima accesare: 10.X.2022.
- Sârbu, R. 1987. *Colocație și coligație în sintagmatica lexico-sintactică*, in „Filologie XXX”, Timișoara, Tipografia Universității din Timișoara, p. 53-64.
- Snow, C.E., Ucelli, P. 2009. *The Challenge of Academic Language*, in vol N. Torrance, D.R. Olson (Eds), *The Cambridge Handbook of Literacy*, Cambridge: Cambridge University Press, p. 112-133.
- Townsend, D., Filippini, A., Collins, P., Biancarosa, G. 2009. *Evidence for the Importance of Academic Word Knowledge for the Academic Achievement of Diverse Middle School Students*, in "The Elementary School Journal", Vol. 112, Nr. 3, p. 497-518.
- Tucan, D., Pop, A.-M., Bercuci, L., Chitez, M. 2020a. *Writing Transition: A Corpus-Supported Study of the Freshmen's Written Argumentation Competence Building*, in "Conference Proceedings. 13th International Conference Innovation in Language Learning". Florence: Filodiritto Editore, disponibil online la <https://conference.pixel-online.net/ICT4LL/files/ict4ll/ed0013/FP/3177-CUD4854-FP-ICT4LL13.pdf>, ultima accesare: 10.X.2022.
- Tucan, D., Rogobete, R., Chitez, M., Radu-Pop, A.-M. 2020b. *Cât de pregătiți sunt elevii de liceu pentru scrierea academică de nivel universitar? Studiu didactic contrastiv bazat pe date de corpus lingvistic*, in „Analele Universității de Vest din Timișoara. Seria Științe Filologice”, vol. LVIII, Editura Universității de Vest, p. 69-92.
- Țeghiu, L. 2004. *Structuri stabile în limba română*. Timișoara: Editura Marineasa.

## Webografie

- Chitez, M., Bercuci, L., Dincă, A., Rogobete, R., Csürös, K. 2021. *Corpus of Romanian Academic Genres* (ROGER). Universitatea de Vest din Timișoara. Disponibil la: <https://roger-corpus.org/>.
- Chitez, M., Rogobete, R., Mureșan, V., Dincă, A. 2022. *Corpus of Expert Writing in Romanian and English* (EXPRES). Universitatea de Vest din Timișoara. Disponibil la: <https://expres-corpus.org/>.  
<https://dacre.projects.uvt.ro/>, ultima actualizare la 30 noiembrie 2022.  
<https://www.phrasebank.manchester.ac.uk/>  
<https://www.sketchengine.eu/british-academic-spoken-english-corpus/>.
- <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>, ultima actualizare la 10 septembrie 2019.  
<https://www.sketchengine.eu/picae-corpus/>.  
<https://www.sketchengine.eu/oxford-corpus-of-academic-english/>.  
<https://www.sketchengine.eu/british-academic-written-english-corpus/>.  
<https://www.sketchengine.eu/cambridge-academic-english-corpus/>.  
<https://www.sketchengine.eu/cambridge-academic-english-corpus/>.
- UEFISCDI. 2022. *Reviste românești indexate/cotate „ISI”*, disponibil online la <https://uefiscdi.gov.ro/resource-862463-rev.rom.isi.15.07.2022.pdf>, ultima accesare: 10.X.2022.  
<https://writing.unitbv.ro/stilul-academic/>.