

**ACADEMIA ROMÂNĂ**

INSTITUTUL DE FILOLOGIE ROMÂNĂ „ALEXANDRU PHILIPPIDE”, IAȘI  
INSTITUTUL DE LINGVISTICĂ „IORGU IORDAN – ALEXANDRU ROSETTI”, BUCUREȘTI  
INSTITUTUL DE LINGVISTICĂ ȘI ISTORIE LITERARĂ „SEXTIL PUȘCARIU”, CLUJ-NAPOCA

**LEXICOGRAFIA ACADEMICĂ ROMÂNEASCĂ.  
PROVOCĂRILE INFORMATIZĂRII  
AcadLexi 2025**

Volumul cuprinde lucrări prezentate la ediția a XIV-a a Colocviului cu participare internațională „Lexicografia academică românească. Provocările informatizării”. AcadLexi 2025 (<https://philippide.ro/acadlexi.html>).

**Coordonatori de volum:** Elena Isabelle Tamba, Mariana Nastasia

**Editori:**

Alexandru Anghelina  
Iulia Barbu-Comaromi  
Călina Ciocoiu  
Alexandra Ilie  
Mircea Minică  
Andreea-Nora Pușcaș

**Referenți științifici:**

Conf. univ. dr. Ana-Maria Minuț  
CS I dr. Gabriela Haja

**Coperta:** Lucian Săcrieru-Dragomir

**ISSN: 3008-6159**

**ISSN-L: 3008-6159**

**DOI:** [https://doi.org/10.52257/30086159\\_2025](https://doi.org/10.52257/30086159_2025)

© 2025 Editorii volumului. Toate drepturile rezervate.

Universitatea Babeș-Bolyai

Presa Universitară Clujeană

Str. Hasdeu nr. 51

400371 Cluj-Napoca, România

Tel./fax: (+40)-264-597.401

E-mail: [editura@ubbcluj.ro](mailto:editura@ubbcluj.ro)

<http://www.editura.ubbcluj.ro/>



## **ACADEMIA ROMÂNĂ**

INSTITUTUL DE FILOLOGIE ROMÂNĂ „ALEXANDRU PHILIPPIDE”, IAȘI  
INSTITUTUL DE LINGVICĂ „IORGU IORDAN – ALEXANDRU ROSETTI”, BUCUREȘTI  
INSTITUTUL DE LINGVICĂ ȘI ISTORIE LITERARĂ „SEXTIL PUȘCARIU”, CLUJ-NAPOCA

# **ACTELE COLOCVIULUI INTERNAȚIONAL „LEXICOGRAFIA ACADEMICĂ ROMÂNEASCĂ. PROVOCĂRILE INFORMATIZĂRII” AcadLexi 2025**

PRESA UNIVERSITARĂ CLUJEANĂ  
2025



# ASPECTE PRIVIND DIGITALIZAREA DEX-ULUI CU EDITORUL LEXICOGRAFIC TLEX

ANA-MARIA BARBU\*, ALEXANDRU ANGHELINA\*\*, MIHAELA MARIN\*\*

## 1. Introducere

Când vorbim de digitalizarea unui dicționar nu ne mai referim la simpla transpunere în format electronic a acestuia, ci ne referim la o aplicație informatică care pune la dispoziție consultarea lui prin criterii de căutare, la existența unui sistem de adnotare a informației, la construirea unei baze de date interogabile și la interfața de interogare propriu-zisă. Se pot desprinde astfel două etape majore în acest demers: prima – redactarea în format electronic a intrărilor de dicționar de către lexicografi și a doua – construirea de către informaticieni a unui sistem informatic de exploatare a conținutului redactat. Etapa de redactare a fost înlocuită în unele situații de una de scanare a dicționarelor după forma lor tipărită.

Pentru ca sistemul informatic să poată funcționa, diferitele tipuri de informație dintr-o intrare de dicționar (*i.e.* lema/cuvânt-titlu, categoria gramaticală, sensul, definiția, informația de uz etc.), dar nu neapărat toate, trebuie să fie delimitate corespunzător prin intermediul unor etichete/adnotări sau reprezentări atribut-valoare, fapt care se realizează în mod curent prin reprezentarea XML (eXtensible Markup Language). Această etapă nu poate lipsi din procesul de informatizare.

În lexicografia românească preocuparea pentru digitalizarea lucrărilor a dat deja rezultate notabile, pe câteva dintre ele le vom prezenta în secțiunea următoare, iar articolul de față propune mai multe elemente de noutate pe linia digitalizării lexicografice. În primul rând, aduce în discuție digitalizarea așteptată a unei lucrări de referință cu mare impact public: *Dicționarul explicativ al limbii române* (DEX<sup>3</sup>), ediția a treia. Ca eforturi anterioare în acest sens, menționăm includerea primelor ediții ale DEX-ului în platforma de lucrări lexicografice *dexonline*, precum și lucrările care vizează digitalizarea DEX<sup>3</sup> la nivel conceptual (Oprea *et al.* 2021, Radu-Bejenaru 2021).

---

\* Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti”, Academia Română, București, Casa Academiei, Calea 13 Septembrie, nr. 13/ Facultatea de Litere, Universitatea din București, str. Edgar Quinet, nr. 5–7.

\*\* Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti”, Academia Română, București, Calea 13 Septembrie, nr. 13.

\*\* Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti”, Academia Română, București, Calea 13 Septembrie, nr. 13.

În al doilea rând, lucrarea aceasta propune folosirea unui editor lexicografic profesional, cu numeroase facilități specializate, care să înlocuiască popularul, dar prea generalul editor al companiei Microsoft: MS Word. Iar în al treilea rând, la fel de important, propune ca în demersul de digitalizare lexicografică să preia etapa importantă a adnotării XML, transferând către informaticieni doar construirea interfeței de interogare. Acest demers, în ciuda unui efort puțin mai mare din partea lexicografului, conferă muncii acestuia mult mai multă flexibilitate și aduce costuri de digitalizare semnificativ mai mici. Un dicționar cu reprezentare XML robust construită de un editor profesional crește foarte mult valoarea dicționarului și acoperă o etapă foarte importantă din demersul de digitalizare.

Prezentul articol conține, așadar, o secțiune, a doua, dedicată unor dicționare de referință deja digitalizate pentru a pune în evidență ce metode s-au folosit și ce metodă nouă se aduce în discuție. În secțiunea principală, a treia, sunt prezentate caracteristicile editorului lexicografic avut în vedere, numit TLex, un exemplu de structurare generală a intrării din DEX cu referire doar la informația de sens și modul general de lucru cu TLex, opțiuni de digitalizare și probleme depistate în etapa de testare. Secțiunea a patra propune o comparație, cu avantaje și dezavantaje, între digitalizarea cu TLex și cea folosită la dicționarele DELR și DOOM<sup>3</sup> care presupune utilizarea editorului MS Word și a unui analizor. Lucrarea se încheie cu secțiunea de concluzii.

## **2. Metode de digitalizare aplicate anterior, pentru DLR, DELR și DOOM<sup>3</sup>**

Pentru a pune în evidență mai bine ce aduce nou digitalizarea cu ajutorul editorului lexicografic TLex, considerăm necesară o trecere în revistă a rezultatelor de digitalizare a unor dicționare de referință, care sunt deja puse la dispoziția publicului larg, anume *Dicționarul-tezaur al limbii române* (DLR), *Dicționarul etimologic al limbii române* (DELR) și *Dicționarul ortografic, ortoepic și morfologic* (DOOM<sup>3</sup>).

### **2.1. DLR și DLRI**

DLR poate fi consultat public pe platforma CLRE la adresa <https://dlr1.solirom.ro/>. Interfața de interogare folosește varianta anastatică a volumelor DA/DLR din care a fost extras fiecare cuvânt-titlu (sau leamnă) și pagina sau paginile pe care se întinde intrarea respectivă, precum și categoria lui gramaticală. Căutările în dicționar se pot face, așadar, după leamnă, ca în fig. 1, în care este ilustrată intrarea *face*. După cum se vede, această intrare se întinde de la pagina 4 (a fasciculei) până la 23, fiecare pagină putând fi accesată prin butonul corespunzător.



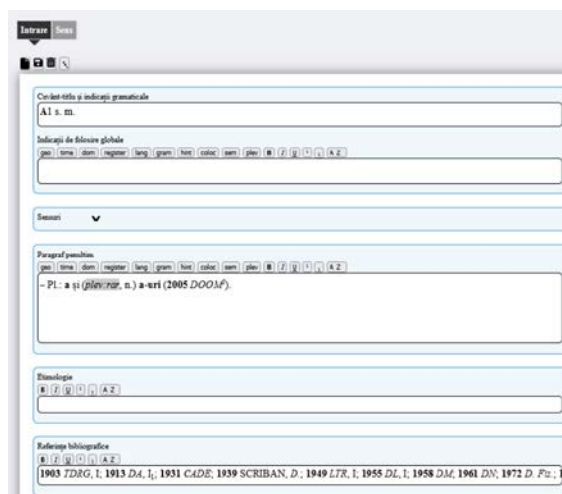


Figura 3. Fereastra de lucru a DLRI.

În urma introducerii datelor de către lexicografi se obține formatul XML, exemplificat în fig. 4:

```

40      </dictScrap>
41    </sense>
42  <sense xml:id="dh8">
43    <def><hi rend="bold">2.</hi> <usg type="sem">P. gener.</usg> <usg type="geo">Reg.</usg> Croitor.</def>
44    <dictScrap>
45      <cit><hi rend="bold">&lt;1929-1938&gt;</hi> <hi rend="italic">ALR</hi> <hi rend="italic">(SN</hi> <hi rend="italic">
46    </dictScrap>
47  </sense>
48  </dictScrap>
49  <form type="varia">
50    <form>- Pl.: <hi rend="bold">abagii</hi></form>
51    <form>- Var.: <usg type="time">Inv.</usg>, <usg type="plev">rar</usg> <hi rend="bold">abăgiu</hi>, <hi rend="bold">abagău</hi>, <hi r
52  </form>
53  <etym>- Din tc. <hi rend="bold">abağyc</hi>.</etym>
54  <note type="bibl-ref"><hi rend="bold">1931</hi> <hi rend="italic">CADE</hi>; <hi rend="bold">1966</hi> CIORĂNESCU, <hi rend="italic">D.</hi>
55 </entry>
56 <note type="persona">
57   <editor role="author"/>
58   <editor role="reviewer"/>
59 </note>
60 <note type="varia">
61   <idno type="lexicon"/>
62   <idno type="entry-status">1</idno>
63 </note>

```

Figura 4. Exemplu de reprezentare XML a DLRI.

Interfața a fost concepută pentru accesare de la distanță cu utilizatori multipli. Această facilitate a dus însă la numeroase probleme de conexiune și la situații de corupere a bazei de date XML din spatele interfeței. La acestea se adaugă faptul că lexicografii își redactau intrările întâi în MS Word, apoi introduceau datele în interfață și că unele dintre elementele de stil din MS Word erau problematice în interfață. De aceea, s-a luat decizia ca interfața DLRI să fie abandonată și s-a adoptat soluția construirii unui analizor (engl. *parser*) al intrărilor redactate în MS Word care să obțină reprezentarea XML și baza de date aferentă și mai departe construirea unei interfețe de interogare. Analizorul stabilește tipurile de informații pe baza mărcilor formale din intrarea de dicționar, mai precis pe baza particularităților de stil (bold, italic, majuscule etc.) folosite în MS Word și a ordonării informației în general, oferite de lexicograf (Mititelu, Anghelina 2025). Această metodă a fost folosită, de asemenea, pentru digitalizarea dicționarelor DOOM<sup>3</sup> și DELR pe care le prezentăm în ceea ce urmează.

## 2.2. DELR și DOOM<sup>3</sup>

Dicționarele DELR și DOOM<sup>3</sup> sunt digitalizate de Cătălin Mititelu (Mititelu 2024) pornind de la fișierele (destinate tipăririi) în format MS Word ale celor două dicționare. Acestea sunt analizate de un program (engl. *parser*), care identifică în intrarea de dicționar informația relevantă pentru adnotare și indexare, această informație putând reprezenta pentru interfața de interogare un criteriu de căutare. De pildă, în DELR (în fig. 5) analizorul a identificat cuvântul-titlu și cuvintele din familia lexicală a acestuia, acestea constituind criteriile de căutare în dicționarul digitalizat. În exemplul din fig. 5 a fost căutat cuvântul *biker*, care face parte din familia lexicală a lemei *bicicletă*.



Figura 5. Digitalizarea DELR.

Digitalizarea DOOM<sup>3</sup> a presupus o analizare mai complexă, criteriile de căutare referindu-se la leme și forme flexionare, dar la acestea sunt adăugate filtre după categoria gramaticală, limbă de proveniență etc., precum și două facilități foarte utile: căutarea aproximativă (care presupune găsirea și a cuvintelor din dicționar care diferă cu puțin de forma căutată, considerată greșită) și folosirea wildcardurilor, adică simboluri care pot înlocui secvențe de litere, cum sunt \*, care înlocuiește oricâte caractere din cuvânt sau ?, care înlocuiește o singură literă. De pildă, în fig. 6 a) sunt căutate toate cuvintele care se termină cu sufixul *-ișor*, iar acestea pot fi filtrate după criteriile din coloana din stânga. Este de notat că aceste filtre pot deveni ele însele criterii de căutare, dacă mai întâi se caută toate cuvintele din DOOM cu \*, apoi se aplică filtrul dorit. Spre exemplu în fig. 6b) au fost căutate toate cuvintele, apoi s-a aplicat filtrul „limba de proveniență: portugheză”.

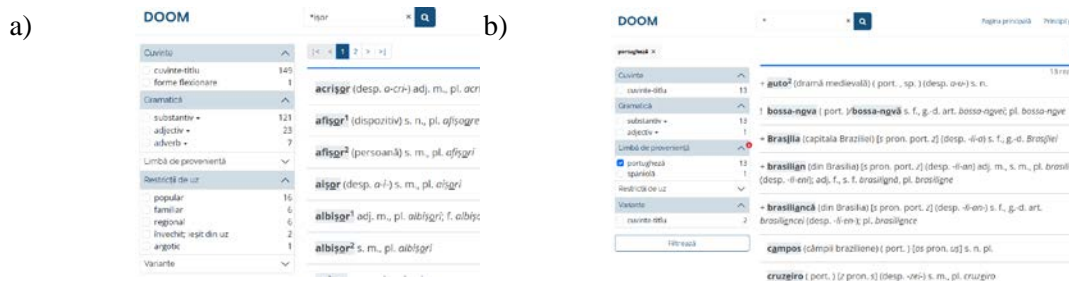


Figura 6. Criterii de căutare în DOOM<sup>3</sup>.

În cazul dicționarelor prezentate mai sus, rolul informaticienilor este esențial în dezvoltarea întregului sistem, lexicografii rămânând la metoda tradițională de redactare a

dicționarelor și, într-o proporție mult mai mică, oferind asistență informaticienilor. În ceea ce urmează propunem, în schimb, o metodă care sporește rolul lexicografilor în digitalizare.

### **3. Digitalizarea DEX-ului cu editorul lexicografic profesional TLex**

Dacă metodele de digitalizare prezentate mai sus pornesc de la dicționare redactate de lexicografi în formatul MS Word, în ceea ce urmează propunem redactarea cu ajutorul unui editor lexicografic profesional, numit TLex, subliniind facilitățile oferite, dar și eventuale limitări.

#### **3.1. Caracteristicile editorului lexicografic TLex**

Editorul TLex este o aplicație informatică, care face parte dintr-un pachet de aplicații dedicat cercetării lexicografice, elaborat și comercializat de o firmă din Africa de Sud, și care poate fi achiziționat pentru o sumă accesibilă (de circa 200 euro) de pe site-ul <https://tshwanedje.com/>.

După cum reiese de pe site-ul companiei, acest software este folosit în edituri și în institute de pe tot cuprinsul mapamondului, dintre care remarcăm edituri de prestigiu precum Oxford University Press, Longman, Macmillan, Le Robert sau institute ca the Royal Spanish Academy, the Institute for Dutch Lexicology, the Malaysian Institute of Language and Literature, South African National Lexicography Units și altele.

Editorul a fost achiziționat și de Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti” și a fost utilizat cu succes în elaborarea lucrărilor *Dicționarul de contexte verbale* (DCV), *Inventarul de cuvinte și sensuri noi atestate în mediul online* (ICSO), volumele 2 și 3, sau *Dicționarul graiurilor românești din Ungaria* (DGRU).

Caracteristica de bază a acestui editor este dată de elaborarea lucrărilor lexicografice în formatul de adnotare XML, ceea ce presupune reprezentarea de tip atribut-valoare, unde atributul este de fapt numele etichetei cu care se adnotează diversele informații lexicografice concrete, vizibile în intrarea de dicționar, văzute ca *valori* ale atributului/etichetei respectiv(e). Adnotarea se face după o structură ierarhică, numită și schemă de adnotare, descrisă în așa-numitul DTD (engl. *Document Type Definition*). Primul pas în utilizarea editorului TLex este, așadar, construirea DTD-ului, după ce a fost stabilită conceptual structura de adnotare a dicționarului (v. secțiunea 3.2). O facilitate esențială este dată de faptul că lexicograful poate construi singur acest DTD, putând opta pentru structuri de dicționar dintre cele mai diverse, a se vedea în acest sens DCV (<https://dcv.lingv.ro/>) care are o structură complet diferită de cea a unui dicționar explicativ sau bilingv obișnuit.

O altă caracteristică a editorului este că odată cu DTD-ul se poate elabora și stilul de tehnoredactare a diferitelor componente ale intrării de dicționar astfel încât să se obțină direct formatul tipăribil. Stilul de prezentare a unei intrări poate fi modificat oricând, chiar și după elaborarea completă a dicționarului. Preluarea de către TLex a stilului degrevează lexicograful de o muncă destul de laborioasă, lăsându-l să se concentreze asupra conținutului. Legat de stil, trebuie să subliniem un aspect important. În prezent, deși ne aflăm

în era digitală, se acordă încă importanță formei tipărite, pentru care TLex oferă numeroase facilități. Acest lucru are însă un dezavantaj, și anume faptul că opțiunile de stil influențează structura DTD-ului complicându-l. Mai precis, uneori trebuie să se includă în structură anumite elemente nu din necesitatea adnotării unui anumit tip de informație, ci din necesitatea adoptării unui anumit stil. Cu alte cuvinte, dacă TLex ar fi folosit exclusiv pentru construirea unui dicționar sub forma unei baze de date, atunci structura DTD ar fi semnificativ mai simplă decât dacă se dorește un dicționar și în format tipărit.

Existența DTD-ului asigură posibilitatea ca o intrare de dicționar să fie elaborată pe elemente nu neapărat într-o ordine secvențială ca în MS Word. De pildă, după cuvântul-titlu se poate scrie etimologia și apoi informațiile gramaticale, toate acestea fiind puse automat în ordinea prestabilită a intrării. Informațiile de același tip, cum ar fi sensurile, pot fi rearanjate ulterior într-o anumită ordine.

TLex permite construirea de baze de date care pot fi accesate de la distanță de către mai mulți utilizatori în același timp. Trebuie menționat însă că uneori accesul de la distanță poate produce coruperi ale bazei (fapt menționat și la DLRI). Din experiența proprie putem spune că dacă pentru baza de date aferentă volumului ICSO 2 nu am avut niciun fel de probleme, pentru volumul 3 am întâmpinat dificultăți, probabil din cauza unor incompatibilități între TLex și dispozitivele lexicografilor conectați la bază.

Dintre numeroasele facilități ale acestui editor, ne limităm a mai aminti caracterul independent de limbă, de pildă, prin folosirea reprezentării Unicode a caracterelor și prin posibilitatea de a defini ordinea alfabetică specifică fiecărei limbi. De asemenea, TLex permite exportarea dicționarelor în diferite formate, cum ar fi RTF (compatibil cu MS Word, pentru tipărire), XML (pentru prelucrări informatice ulterioare), HTML (pentru afișarea pe internet) etc. Pentru ultimele variante ale TLex, producătorii anunță posibilitatea integrării ChatGPT în elaborarea dicționarelor.

### 3.2. Exemplu de structurare pentru DEX și modul de lucru cu TLex

Pentru ediția în lucru a DEX-ului au fost preluate informațiile de conținut și stil din Moroianu 2021. Acestea au fost configurate în DTD-ul editorului lexicografic, care folosește formatul XML. După cum se știe, în acest format, se lucrează cu două componente de bază: *elemente* și *attribute*. Elementele sunt componentele cu structură internă, asemenea ramurilor unui arbore, iar attributele sunt componentele terminale, nestructurate, asemenea frunzelor, cărora, în planul concret al unui dicționar, le corespund fragmente de text din intrarea de dicționar. În structura internă a unui element pot intra alte elemente, numite *copii*, și/sau attribute. În fig. 7 este reprezentată structura conceptuală a unui sens din DEX<sup>3</sup>. Attributele sunt cele care conțin indicația *text*, pentru locul în care lexicograful va introduce textul concret din intrare, iar restul reprezintă elemente pe care lexicograful le va selecta pentru a construi structura de adnotare (invizibilă) adecvată unei anumite intrări.

Sens	NivelSens: text	DefText: text	
	Cat.Gram: text		Exemple: text
	IndicatieUZ: text		
Definiție	DefNorm	IndicatieUZ: text	
	SintExprEtc.	[...]	
	Sinonime	[...]	
	TrimitereV.	[...]	

Figura 7. Structura elementului Sens.

După cum se observă, elementul Sens este alcătuit din 3 atribute: NivelSens în care se scrie numărul aferent sensului, potrivit filiației sensurilor; Cat.Gram – categoria gramaticală aferentă lemei (e.g. s.m., adj., vb.tranz. etc.), IndicatieUZ – informații de domeniu, timp, frecvență, context etc.; și un alt element, Definiție, care la rândul lui poate avea drept copil o definiție normală (DefNorm), o definiție de tip sintagmă, locuțiune, expresie etc. (SintExprEtc.), o definiție de tip sinonime (Sinonime) sau o trimitere la altă intrare (TrimitereV). Definiția normală (DefNorm) poate fi alcătuită din atributele în care se scrie definiția propriu-zisă (DefText), exemple (Exemple) și/sau indicații de uz (IndicatieUZ).

O structură conceptuală, ca cea din fig. 7, care să acopere toate structurile posibile a intrărilor din DEX a fost transpusă în DTD-ul editorului Tlex și constituie baza de lucru în maniera prezentată în ceea ce urmează.

În fig. 8 este redată fereastra de lucru a editorului lexicografic. Aceasta conține 4 sub-ferestre: cea din stânga redă alfabetic inventarul de cuvinte introduse, cea din mijloc – sus servește la structurarea intrării prin alegerea adecvată a elementelor, cea din mijloc – jos conține atributele, acestea urmează să fie completate cu text (atributele care nu sunt completate cu text rămân în structură, dar sunt ignorate de aplicație), iar fereastra din dreapta previzualizează textul introdus potrivit stilului specificat în DTD.

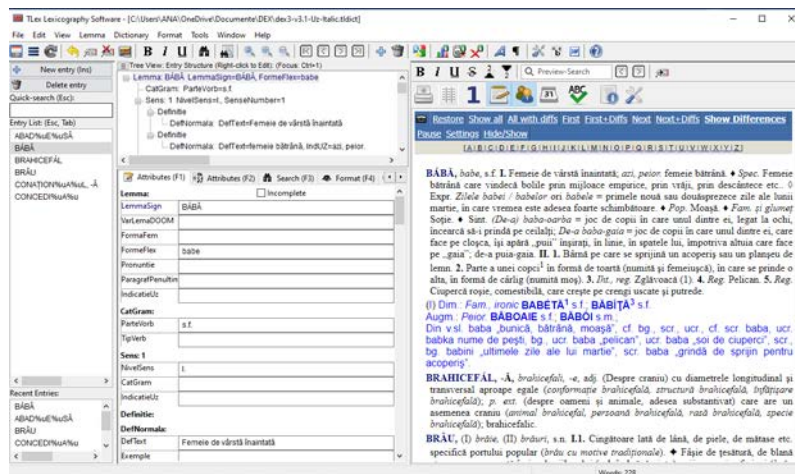


Figura 8. Fereastra de lucru a Tlex.

Lexicograful începe lucrul prin introducerea unei noi intrări (butonul *New entry* din stânga sus în fig. 8) și completarea lemei în câmpul care apare. Apoi completează textul atributelor în funcție de ce elemente adaugă la structura intrării. De pildă, în fig. 9 sunt prezentate atributele Lemei și corespondentele lor din intrarea de dicționar.

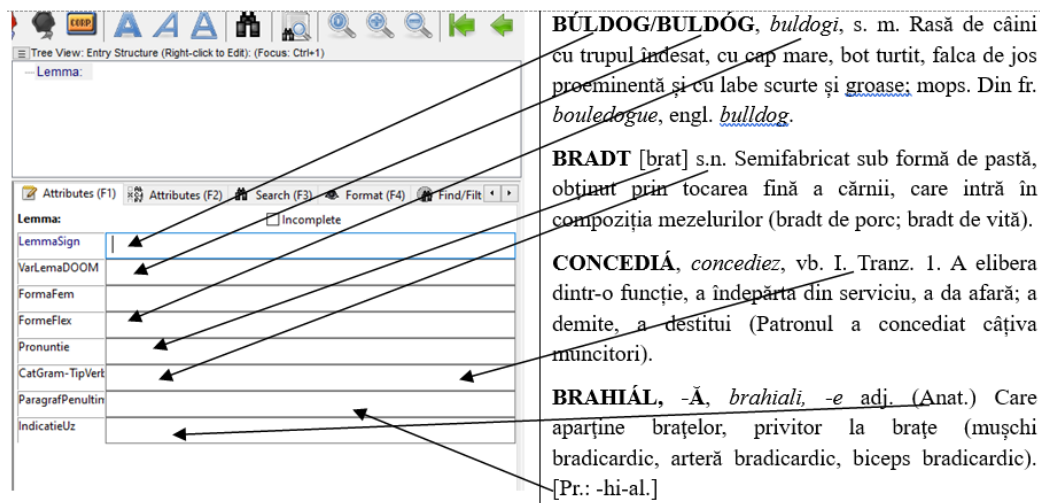


Figura 9. Atributele Lemei.

În continuare, lexicograful alege copiii Lemei, din opțiunile prezentate în fig. 10. Aceștia pot fi sensuri (Sens), una sau mai multe variante speciale (învechite sau regionale: VarSpecSet), un set de una sau mai multe leme asociate (LemeAsociateSet), etimologia (Etimologie) și o trimitere la altă intrare (TrimitereV).

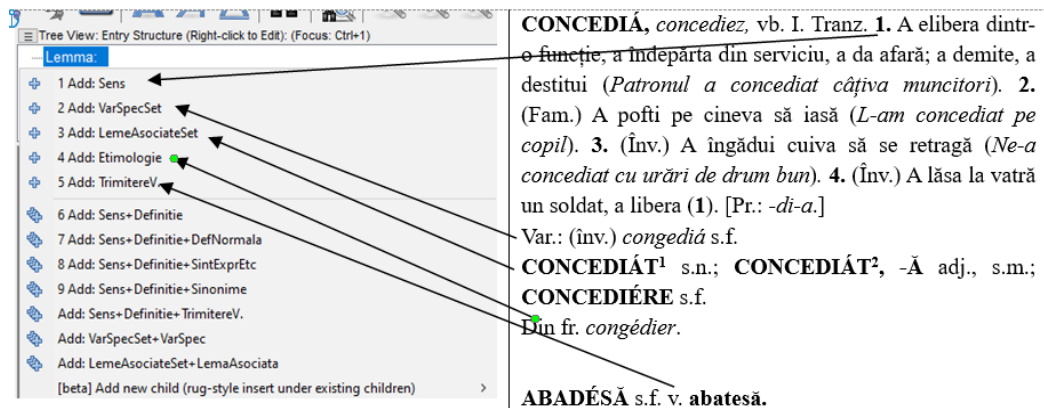


Figura 10. Copiii Lemei.

Este de notat că TLex pune la dispoziția utilizatorilor scurtături structurale, adică secvențe de copii, după cum se vede în fig. 10 în stânga de la punctul 6 încolo, de pildă prin Sens+Definitie+DefNormala se introduc direct toate elementele din calea respectivă.

### 3.3. Soluții de compromis în digitalizarea DEX-ului

În demersul de digitalizare a DEX-ului cu editorul lexicografic TLex am adoptat o serie de opțiuni care au ca motivație, în primul rând, simplificarea muncii lexicografilor, etichetând, în principiu, doar acel tip de informație care este relevant pentru căutări sau pentru alte aplicații informatice, în așa fel încât structurarea intrărilor de către lexicografi să nu devină prea laborioasă.

În acest sens, am optat pentru nume de etichete (aferele elementelor și atributelor) cât mai sugestive, în limba română, renunțând la standardul TEI. Am renunțat de asemenea la ierarhizarea internă a sensurilor, la imbricarea lor în reprezentarea XML, altfel spus, toate sensurile au același nivel de imbricare, anume drept copii ai lemei (nu copii ai altor sensuri). Marcarea filiației sensurilor, despre care știm din experiență că este imposibil de realizat automat, se face doar prin atributul NivelSens, care poate avea valori de tipul A, B, etc., I, II, etc. 1, 2, etc., a), b) etc. (sau combinații ale acestora), ◆ sau ◇. Pe baza acestor valori se poate reconstitui ierarhizarea sensurilor, dacă aceasta este relevantă pentru anumite cercetări, printr-o aplicație informatică specializată.

Tot pentru simplificarea schemei de adnotare, nu etichetăm în detaliu toate tipurile de informații, de pildă, câmpul CatGram include informații de gen, număr, conjugare etc., fără să alocăm câte un atribut pentru fiecare dintre acestea. Un câmp similar este alocat Etimologiei, care include toate informațiile etimologice atât de diferite de la o intrare la alta, fără să etichetăm etimonul, ci doar limbile-sursă și tipul de etimologie. Tot câmp cuprinzător este și cel al indicațiilor de uz (IndicatieUz), care include domeniu, frecvență, arie geografică etc. Considerăm că eventuale căutări din interfața de interogare se pot face după cuvinte cheie de tipul *med.*, *bot.* etc., sau *pop.*, *colocv.*, *reg.* etc.

Pe de altă parte, nu folosim referințe încrucișate, adică trimiteri dintr-o intrare la alte intrări sau sensuri, deși TLex le permite, tot din dorința de a simplifica lucrul în această fază de început în utilizarea editorului lexicografic.

### 3.4. Probleme și soluții în utilizarea TLex-ului

Schema de adnotare a următoarei ediții a DEX-ului a fost testată pentru circa 150 de intrări de dicționar (redactate în TLex de Mihaela Marin și Ștefania Oprea). Problemele semnalate sunt legate de faptul că TLex include, cum menționam anterior, elemente de stil. Acestea sunt aplicate automat și pot intra în conflict cu textul propriu-zis. În ceea ce urmează exemplificăm o astfel de problemă legată de faptul că TLex pune automat punct final după elementul Sens, LemeAsociate și Lemma (la sfârșitul întregii intrări de dicționar). După cum se vede în fig. 11a), dacă există un sens în care este folosit doar atributul NivelSens și în rest nu conține nicio altă informație, atunci apar două puncte separate prin spațiu (încercuite cu roșu în figură).

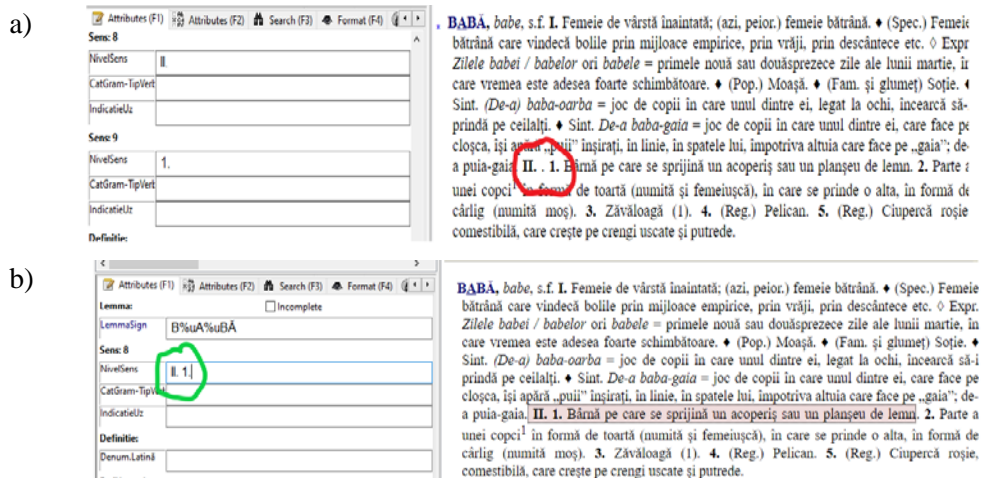


Figura 11. Problema punctului după sens gol.

Soluția este să se renunțe la sensul gol și să se comaseze cele două valori în atributul NivelSens, ca în fig. 11b) (în zona marcată cu verde).

Aceeași problemă este întâlnită când un sens – sau un câmp în general –, la care se adaugă automat punct final, se termină cu o abreviere. În această situație apar iarăși două puncte consecutive, ca în fig. 12a). Soluția este ca lexicograful să șteargă punctul abrevierii, atunci când vede în fereastra de previzualizare a TLex-ului că apare această problemă, ca în fig. 12b).

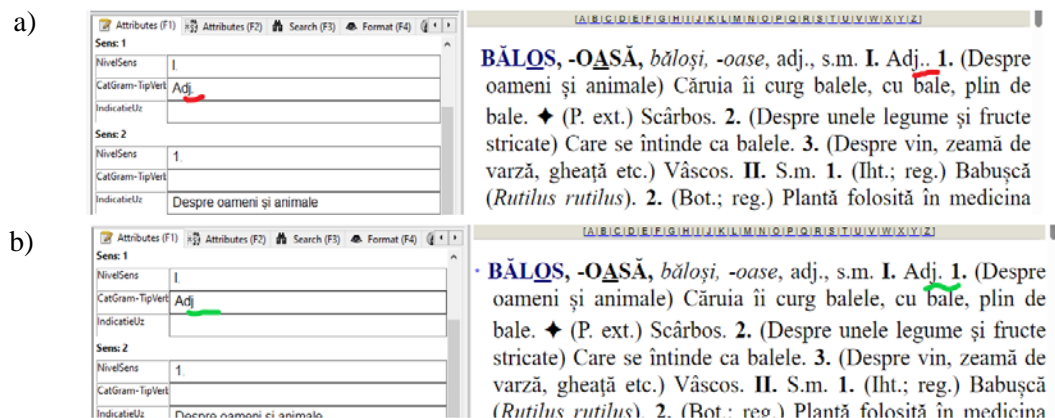


Figura 12. Problema punctului după o abreviere.

#### 4. Comparație între digitalizarea cu TLex și metoda de digitalizare Word/analizor

După cum am arătat în secțiunea 2, lexicografii redactează dicționarele mari aproape exclusiv în MS Word pentru a putea fi tipărite pe hârtie, iar procesul de digitalizare, care reprezintă o etapă ulterioară, fie dublează această muncă (DLRi), fie se bazează formal pe ea (DOOM<sup>3</sup>, DELR).

Din punctul de vedere al tehnoredactării, TLex, prin faptul că este un editor special conceput pentru lucrul lexicografic, prezintă mai multe avantaje față de MS Word, dintre care le menționăm pe următoarele:

- informația introdusă este pusă în locul prestabilit din intrare indiferent de ordinea redactării;
- există numeroase operații care se fac automat (stiluri, paranteze, puncte finale etc.);
- preluarea automată a stilului permite lexicografului să se concentreze pe conținut;
- ordonarea alfabetică a intrărilor și numerotarea omonimelor se face automat;
- există multe alte facilități lexicografice (baze de date online, descrieri multimedia, referințe încrucișate, legături cu exemple din corpus etc.).

În privința dezavantajelor TLex-ului, față de MS Word, putem menționa:

- nu are bogăția de fonturi a MS Word, de pildă, nu poate spația caracterele unui cuvânt așa cum s-a practicat în edițiile anterioare ale DEX-ului pentru reprezentarea etimoanelor;
- nu are un corector ortografic pentru limba română, deși sunt incluse corectoare pentru alte limbi. Este posibil ca acest dezavantaj să poată fi rezolvat printr-o cercetare dedicată acestui subiect;
- nu are funcția „Split windows”, care este utilă pentru anumite operații;
- prezintă unele probleme ca cele redate în secțiunea 3.4.

Din punctul de vedere al digitalizării, TLex prezintă, în opinia noastră, următoarele avantaje față de utilizarea analizorului de text (ca pentru DELR și DOOM<sup>3</sup>). Structurarea intrării (*i.e.* schema de adnotare) este complet accesibilă lexicografului, în timp ce analizorul este construit exclusiv de un informatician. TLex este independent de tipul dicționarului, lexicograful putând să-și configureze intrări dintre cele mai diverse, în timp ce analizorul este strict dependent de tipul de dicționar și de normele de tehnoredactare ale dicționarului analizat. Dacă aceste norme de redactare nu sunt strict respectate, ceea ce în practică se întâmplă foarte frecvent, pot exista erori mari în construirea fișierului XML, mai precis, în delimitarea diferitelor tipuri de informații din intrare. De aceea, este absolut obligatoriu un efort de verificare minuțioasă și completă a rezultatelor analizorului. În schimb, în TLex erorile de tehnoredactare nu afectează structura XML, rămânând limitate în interiorul câmpului completat de lexicograf.

Ca un dezavantaj pentru TLex se poate considera faptul că parte din digitalizarea dicționarului este preluată de lexicograf, el trebuind să preia efortul de structurare a intrării curente. La aceasta se poate adăuga și efortul de început, de familiarizare cu această aplicație.

Un alt dezavantaj al TLex, care poate fi considerat major, este că necesită o interfață de interogare care nu poate fi construită decât de un informatician. Dacă se dorește doar elaborarea unui dicționar care să fie tipărit pe hârtie, TLex este suficient, dar digitalizarea nu ar fi completă. În privința interfeței de interogare nu se poate renunța la informatician, iar, practic, informaticianul care construiește analizorul construiește și interfața de interogare într-un produs integrat. Totuși, în opinia noastră, TLex oferă lexicografului mult mai multă

flexibilitate și în ceea ce privește tehnoredactarea, și în ceea ce privește valorificarea ulterioară a formatului XML, care poate fi privit ca o formă de digitalizare de bază, realizată exclusiv de lexicograf. O interfață de interogare construită pe o descriere XML robustă (fără erori), ca cea oferită de TLex, nu ridică nici pe departe atâtea probleme ca un analizor al textului MS Word.

Considerăm că o interfață de interogare ar fi mult simplificată și, eventual, ar putea fi construită cu aplicații de IA, dacă s-ar renunța la obiectivul dicționarului tipărit, iar schema de adnotare a TLex ar fi construită exclusiv pentru conținut, fără stil, ca o simplă bază de date. Aceasta ar putea duce la digitalizarea completă a unui dicționar prin efortul exclusiv al lexicografilor, ajutați, firește, de IA.

## 5. Concluzii

În acest articol am prezentat rezultatele unei etape importante din digitalizarea DEX<sup>3</sup>, care constă în construirea schemei de adnotare XML a dicționarului și implementarea ei într-un editor lexicografic profesional. În urma testării pe circa 150 de intrări, problemele apărute sunt minore și pentru ele s-au identificat soluții. În plus, am arătat avantajele atât pentru tehnoredactare, cât și pentru digitalizare a utilizării editorului TLex prin comparație cu metoda de digitalizare bazată pe MS Word și analizor, dar nu am trecut cu vederea nici dezavantajele acestuia.

Trebuie subliniat că editorul TLex nu oferă o digitalizare completă, fiind necesară munca de informatician pentru construirea interfeței de interogare, dar oferă o flexibilitate mult mai mare în digitalizarea oricărui tip de dicționar, conferă valoare sporită dicționarului creat și reduce mult costurile de digitalizare completă.

Următoarea etapă în digitalizarea DEX-ului este instruirea lexicografilor pentru folosirea editorului TLex și pentru redactarea propriu-zisă a intrărilor.

## BIBLIOGRAFIE

- Anghelina, Teodorescu 2020 = Alexandru Anghelina, Claudiu Teodorescu, *Digitalizarea Dicționarului limbii române. Probleme. Soluții. Perspective*, în „Limba Română”, LXIX, nr. 3–4, p. 491–496.
- Busuioc, Teodorescu 2013 = Monica Busuioc, Claudiu Teodorescu, *Informatizarea Dicționarului limbii române – un vis pe cale de realizare*, în Marius Sala et al. (ed.), *Lucrările celui de-al cincilea Simpozion internațional de lingvistică. București, 27–28 septembrie 2013*, București, Univers Enciclopedic Gold, p. 514–528.
- Haja et al. 2021 = Gabriela Haja, Elena Isabelle Tamba, Marius-Radu Clim, Claudiu Teodorescu, Alexandru Anghelina, *Metode și tehnici actuale în redactarea Dicționarului limbii române informatizat (DLRi)*, în Ofelia Ichim (coord.), *România în spațiul euroatlantic: interferențe culturale și lingvistice*, București, Editura Tracus Arte, 2020, p. 139–160.
- Mititelu 2024 = Cătălin Mititelu, *DOOM<sup>3</sup> digitalization*, prezentare la *Al IX-lea Simpozion Internațional de Lingvistică*, 23–24 mai, Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti”, București.

- Mititelu, Anghelina 2025 = Cătălin Mititelu, Alexandru Anghelina, *O nouă interfață grafică pentru DLR*, în Elena Isabelle Tamba, Mariana Nastasia (coord.), *Actele Colocviului Internațional „Lexicografia academică românească. Provocările informatizării”*. AcadLexi 2025, Cluj-Napoca, Presa Universitară Clujeană, 2025, p. 143–153.
- Moroianu 2021 = Cristian Moroianu, *Dicționarul explicativ al limbii române (DEX<sup>3</sup>)*. *Propuneri structurale*, în „Studii și cercetări lingvistice”, LXXII, nr. 2, p. 165–184.
- Oprea *et al.* 2021 = Ștefania Oprea, Alexandru Anghelina, Claudiu Teodorescu, *Propuneri pentru digitalizarea DEX<sup>3</sup>*, „Studii și cercetări lingvistice”, LXXII, nr. 2, p. 184–194.
- Radu-Bejenaru 2021 = Andreea Radu-Bejenaru, *Modele de digitalizare și aplicabilitatea lor la noua ediție a Dicționarului explicativ al limbii române*, „Studii și cercetări lingvistice”, LXXII, nr. 2, p. 295–305.

## **ASPECTS REGARDING THE DIGITALIZATION OF THE DEX WITH THE TLEX LEXICOGRAPHIC EDITOR**

### **ABSTRACT**

This article presents a new method for digitizing the Explanatory Dictionary of the Romanian Language (DEX<sup>3</sup>) compared to the methods used for the computerization of the DLR, DELR and DOOM<sup>3</sup>. The novelty consists in writing the DEX with the lexicographic editor TLex instead of the general editor MS Word and in the construction by lexicographers of the XML representation of the dictionary. In this regard, the article presents the main features of TLex and the way of operating with this editor, an example of XML structuring of meaning information, notes the problems and solutions that arose during the testing period and makes a comparison between this method and the previous ones.

**Keywords:** *digitalization, DEX, lexicographic editor TLex, XML, query interface, DLR, DELR, DOOM.*

