

GHID DE UTILIZARE

Corpus de română ca limbă străină (LECOR)¹

<http://lecor.unibuc.ro/crystal/#open>

- **Detalii** despre construirea corpusului și despre metadate:

Barbu, Ana-Maria, Elena Irimia, Carmen Mîrzea Vasile, Vasile Păiș, 2023, „Designing the LECOR Learner Corpus for Romanian”, în G. Angelova, M. Kunilovskaya, R. Mitkov (ed.), *Deep Learning for Natural Language Processing Methods and Applications (Proceedings of International Conference Recent Advances in Natural Language Processing, RANLP 2023, Varna, 4–6 September, 2023)*, 143-152 (<https://aclanthology.org/2023.ranlp-1.16.pdf>).

Mîrzea Vasile, Carmen, Elena Irimia, 2023, „Metadata design for the first electronic learner corpus of Romanian”, *Romanian Studies Today VII, 2023*, București, Editura Universității din București, 31-47 (<https://journals.unibuc.ro/index.php/romanian-studies-today/article/view/122>).

- **LECOR** se accesează prin interfața **noSketch engine**. Multe dintre facilitățile oferite de Sketch engine (<https://www.sketchengine.eu/what-can-sketch-engine-do/>) sunt disponibile așadar și pentru LECOR.

- **LECOR** (*Learner Corpus of Romanian*) este un corpus de **română ca interlimbă / ca limbă străină**. Are următoarele **caracteristici**:

- este un corpus monolingv (conține producții ale nenativilor care au studiat limba română ca limbă străină în spațiul vorbitor, în România, în perioada 2019-2024, în mediu academic, într-un program intensiv de un an universitar – cu 28-32 de ore de studiu în clasă pe săptămână, timp de 28 de săptămâni);

- studenții au limbi native diferite (majoritatea orientale și sud-est europene);

- studenții au vârste între 18 și 25 de ani, în general;

- textele colectate sunt de la nivelul A1 până la nivelul B2, câteva și de nivelul C1 (nivelul este evaluat de profesorii lingviști);

- este un corpus alcătuit din texte scrise (de mână sau la calculator) și de transcrieri (ortografice) de producții orale; fiecare text (cu greșeli) al studentului este însoțit de o variantă corectată (cu corecturi minime, de grafie, de gramatică, de vocabular) și de varianta originală a textului / a fișierului audio; toate fișierele au fost anonimizate și pseudonimizate manual;

- este un corpus sincron (conține texte cu subiecte similare produse de mai mulți studenți diferiți în aceleași condiții) și longitudinal / diacronic (conține mai multe texte produse de același student pe parcursul unui an universitar, uneori și peste 100 de texte);

¹ This work was supported by a grant of the Romanian Ministry of Education and Research, CNCS – UEFISCDI, project number PN-III-P1-1.1-TE-2019-1066 (Learner Corpus of Romanian (LECOR). Collection, Annotation and Applications / Corpus de română ca limbă străină (LECOR). Colectare, adnotare și aplicații). Corpusul a fost realizat așadar prin proiectul *Corpus de română ca limbă străină (LECOR). Colectare, adnotare și aplicații* (PN-III-P1-1.1-TE-2019-1066, contract nr. TE 30 / 2020).

- textele sunt adnotate cu metadate detaliate despre student și despre contextul producerii textului și despre tipul de text; data producerii textului permite ordonarea cronologică a textelor;

- textele sunt lematizate și adnotate morfosintactic în platforma RELATE; diferențele dintre cele două variante de text (textul studentului și textul corectat minimal, „al profesorului”) sunt adnotate ca erori (omisiuni, redundanțe, substituție de forme); pentru formele greșite pot fi vizualizate formele corecte;

- este un corpus în acces gratuit; exemplele se pot salva; există acces la textul / fișierul audio original integral.

Variantele LECOR

► **LECOR** este disponibil în **patru variante**, diferite prin tipurile de texte incluse și prin tipul de variante disponibile.

1) LECOR – include textele studenților și varianta lor corectată, iar adnotarea erorilor și a formelor corecte sunt vizibile pe același rând, colorat (v. Fig. 1, cu cea mai simplă vizualizare: forma greșită este cu roșu, forma corectă este cu verde).

| | Details | Left context | KWIC | Right context |
|---|---|--------------|------|---|
| 1 | doc#145 te lucruri pentru distracție . 12 iulie . Astazi Astăzi a fost sămbăta sămbătă . Nu | | am | mers la plaja plajă , am preferăm preferat să stăm |
| 2 | ie . 12 iulie . Astazi Astăzi a fost sămbăta sămbătă . Nu am mers la plaja plajă , | | am | preferăm preferat să stăm puțin la hotel la piscina |
| 3 | plajă , am preferăm preferat să stăm puțin la hotel la piscina piscină , pentru că | | am | vrut să mergem la orașul unsulei insulei seara . La |
| 4 | pentru că am vrut să mergem la orașul unsulei insulei seara . La oră ora 19:00 | | am | mers la orașul oraș cu 2 mașine mașini -0- pe care |
| 5 | La oră ora 19:00 am mers la orașul oraș cu 2 mașine mașini -0- pe care -0- le- | | am | inchiriat inchiriat . Am facut făcut o plimbare și a fi |
| 6 | mers la orașul oraș cu 2 mașine mașini -0- pe care -0- le- am inchiriat inchiriat . | | Am | facut făcut o plimbare și a fost foarte bine . La ora |
| 7 | inchiriat inchiriat . Am facut făcut o plimbare și a fost foarte bine . La ora 22:00 | | am | mers să mâncăm măncăm la o taverna tavernă și |

Figura 1

2) LECOR All texts 2 – include textele studenților și varianta lor corectată, iar adnotarea erorilor și a formelor corecte sunt vizibile pe rândul inferior, cu gri (v. Fig. 2, selectând pentru vizualizare „words” și „teacher”).

| | Details | Left context | KWIC | Right context |
|---|--|--------------|------|--|
| 1 | doc#145 oteluri , și multi lucruri pentru distracție . 12 iulie . Astazi a fost sămbăta . Nu | | am | mers la plaja , am preferăm să stăm puțin la h |
| 2 | lucruri pentru distracție . 12 iulie . Astazi a fost sămbăta . Nu am mers la plaja , | | am | preferăm să stăm puțin la hotel la piscina , per |
| 3 | Nu am mers la plaja , am preferăm să stăm puțin la hotel la piscina , pentru că | | am | vrut să mergem la orașul unsulei seara . La or |
| 4 | la piscina , pentru că am vrut să mergem la orașul unsulei seara . La oră 19:00 | | am | mers la orașul cu 2 mașine -0- care -0- am inc |
| 5 | orașul unsulei seara . La oră 19:00 am mers la orașul cu 2 mașine -0- care -0- | | am | inchiriat . Am facut o plimbare și a fost foarte b |
| 6 | și seara . La oră 19:00 am mers la orașul cu 2 mașine -0- care -0- am inchiriat . | | Am | facut o plimbare și a fost foarte bine . La ora 2 |
| 7 | -0- care -0- am inchiriat . Am facut o plimbare și a fost foarte bine . La ora 22:00 | | am | mers să mâncăm la o taverna și la 23:30 am n |

Figura 2

3) **LECOR Student texts** – include doar textele studenților, cu greșeli, ceea ce permite studii care țin cont strict de numărul de cuvinte al textelor nenativilor (v. Fig. 3).

| | Left context | KWIC | Right context |
|---|---|------|--|
| 1 | doc#145 , hoteluri , și multi lucruri pentru distracție . 12 iulie . Astazi a fost sâmbăta . Nu | am | mers la plaja , am preferăm să stăm puțin la hotel la |
| 2 | doc#145 lucruri pentru distracție . 12 iulie . Astazi a fost sâmbăta . Nu am mers la plaja , | am | preferăm să stăm puțin la hotel la piscina , pentru că |
| 3 | doc#145 Nu am mers la plaja , am preferăm să stăm puțin la hotel la piscina , pentru că | am | vrut să mergem la orașul unsulei seara . La oră 19:00 |
| 4 | doc#145 la piscina , pentru că am vrut să mergem la orașul unsulei seara . La oră 19:00 | am | mers la orașul cu 2 mașine care am inchiriat . Am fac |
| 5 | doc#145 gem la orașul unsulei seara . La oră 19:00 am mers la orașul cu 2 mașine care | am | inchiriat . Am facut o plimbare și a fost foarte bine . Lt |
| 6 | doc#145 unsulei seara . La oră 19:00 am mers la orașul cu 2 mașine care am inchiriat . | Am | facut o plimbare și a fost foarte bine . La ora 22:00 ar |
| 7 | doc#145 așine care am inchiriat . Am facut o plimbare și a fost foarte bine . La ora 22:00 | am | mers să mâncăm la o taverna și la 23:30 am mers la |

Figura 3

4) **LECOR Teacher texts** – include numai textele corectate („ale profesorului”), variante care permit studii mai precise despre formele (cel mai probabil) „intenționate” de student (v. Fig. 4).

| | Left context | KWIC | Right context |
|---|---|------|---|
| 1 | doc#145 , hoteluri și multe lucruri pentru distracție . 12 iulie . Astăzi a fost sâmbătă . Nu | am | mers la plajă , am preferat să stăm puțin la hotel la |
| 2 | doc#145 lucruri pentru distracție . 12 iulie . Astăzi a fost sâmbătă . Nu am mers la plajă , | am | preferat să stăm puțin la hotel la piscină , pentru c |
| 3 | doc#145 . Nu am mers la plajă , am preferat să stăm puțin la hotel la piscină , pentru că | am | vrut să mergem la orașul insulei seara . La ora 19: |
| 4 | doc#145 la piscină , pentru că am vrut să mergem la orașul insulei seara . La ora 19:00 | am | mers la oraș cu 2 mașini pe care le- am închiriat . |
| 5 | doc#145 n la orașul insulei seara . La ora 19:00 am mers la oraș cu 2 mașini pe care le- | am | închiriat . Am făcut o plimbare și a fost foarte bine |
| 6 | doc#145 ulei seara . La ora 19:00 am mers la oraș cu 2 mașini pe care le- am închiriat . | Am | făcut o plimbare și a fost foarte bine . La ora 22:00 |
| 7 | doc#145 ce care le- am închiriat . Am făcut o plimbare și a fost foarte bine . La ora 22:00 | am | mers să mâncăm la o tavernă și la 23:30 am mers |

Figura 4

Pentru toate cele patru variante, de la simbolul pentru vizualizare pot fi selectate tipurile de informații / adnotări care se doresc a fi văzute atunci când se fac căutări.

Exemple de căutări în LECOR

► Din pagina de pornire a LECOR, prima variantă a corpusului prezentată mai sus (v. Fig. 5), se pot iniția căutări în diverse moduri se pot restrânge căutările la anumite tipuri de texte (Fig 6). Explicitarea etichetelor și descrierea metadatelor se găsesc în Mîrzea Vasile, Irimia (2023).

The dashboard is titled 'LECOR' and features a search bar with a magnifying glass icon and an information icon. Below the search bar, there are several tool cards:

- Concordance**: Examples of use in context
- Parallel Concordance**: Translation search
- Wordlist**: Frequency list
- Keywords**: Terminology extraction
- Trends**: Diachronic analysis, neologisms
- Text type analysis**: Statistics of the whole corpus

A 'CORPUS INFO' button is located in the top right corner of the dashboard area.

Figura 5

The 'Text types' section is titled 'Text types ? ^' and includes 'expand all' and 'collapse all' links. It contains a grid of 24 dropdown menus for selecting text types:

- doc.file
- doc.n
- doc.sid
- doc.wid
- doc.age
- doc.gender
- doc.education_level
- doc.native_lang
- doc.bilingual
- doc.proficiency_level
- doc.task_type
- doc.task_requirements_ad_material
- doc.task_mode
- doc.text_function
- doc.particular_pragmatic_function
- doc.topic
- doc.no_of_words_s
- doc.no_of_words_t
- doc.ref_source_usage
- doc.ai_source_usage
- doc.date
- doc.writing_type

A 'GO' button is located at the bottom right of the grid.

Figura 6

În LECOR se pot iniția căutări în moduri diferite:

- 1) Se face click pe opțiunea Concordance, apoi se precizează și / sau se restrâng variantele de texte (Text types) în care se face căutarea.
- 2) Se face click pe opțiunea Text type analysis, apoi se alege un anumit tip de texte în care se face căutarea. Opțiunile de căutări sunt după aceea identice.

1) Pentru primul mod de a iniția căutarea, se alege opțiunea Concordance, apoi se precizează și / sau se restrâng variantele de texte (Text types) în care se face o căutare. Se pot face căutări simple (BASIC) sau avansate (ADVANCED) în tot corpusul sau în anumite tipuri de texte. De exemplu, în Fig. 7, se caută forma „bun” în textele nativilor bulgari și greci. Rezultatele sunt în Fig. 8.

CHANGE CRITERIA

BASIC ADVANCED ABOUT ERROR ANALYSIS

Simple search [?]
bun

Text types (1) [?] [^]

doc.file ▾

doc.n ▾

doc.wid ▾

doc.age ▾

doc.education_level ▾

doc.native_lang [^]
 Bulgarian × Greek ×
 ↔ ▾
 Albanian
 American Spanish

doc.sid

doc.gei

doc.bili

Figura 7

CONCORDANCE

LECOR

Text types 1 (2) ... simple bun • 160
 288.81 per million tokens • 0.029%

Details Left context KWIC Right context

| | | |
|---|--------------------------|--|
| 1 | <input type="checkbox"/> | doc#135 :ocială și bogăție bogăție , toți oameni trebuie să fie educați . -Buna ++ -0 - -0 Bună ziua ! -0 - -Buna Bună ziua , doctor doctore . |
| 2 | <input type="checkbox"/> | doc#135 e , toți oameni trebuie să fie educați . -Buna ++ -0 - -0 Bună ziua ! -0 - -Buna Bună ziua , doctor doctore . -Ce ++ -Ce ++ -0 - -0 |
| 3 | <input type="checkbox"/> | doc#135 :Ce înseamnă -0 că nu am dormit corect ? - Nu ai avut atitudine ++ -0 postură bună . -0 - -Și Și ce trebuie să fac ? -0 - -Trebuie |
| 4 | <input type="checkbox"/> | doc#145 club din insula insulă . Am plecat la 04:30 noaptea . 13 iulie Clubul a fost foarte bun , dar nu vrem să mergem din nou . A fost obo |
| 5 | <input type="checkbox"/> | doc#147 0- pictat . Este o gallery ++ -0 galerie și este foarte frumoș frumos . Nick : Alo , bună ziua . Recepția : Bună ziua . Cum pot să vă a |
| 6 | <input type="checkbox"/> | doc#147 y ++ -0 galerie și este foarte frumoș frumos . Nick : Alo , bună ziua . Recepția : Bună ziua . Cum pot să vă ajut ? Nick : Vreau să fa |
| 7 | <input type="checkbox"/> | doc#148 › , Este pictor și are o compania companie de construcție . Este o un tată foarte bun , el este rabdator răbdător și are sfaturi foarte |
| 8 | <input type="checkbox"/> | doc#148 truție . Este o un tată foarte bun , el este rabdator răbdător și are sfaturi foarte bune . Relația cu el este foarte bun bună , discutan |

Figura 8

Pot fi combinate oricât de multe criterii de limitare a căutărilor. În Fig. 9, la criteriul anterior am adăugat criteriul nivelului de competență, pentru care am selectat valoarea A1.

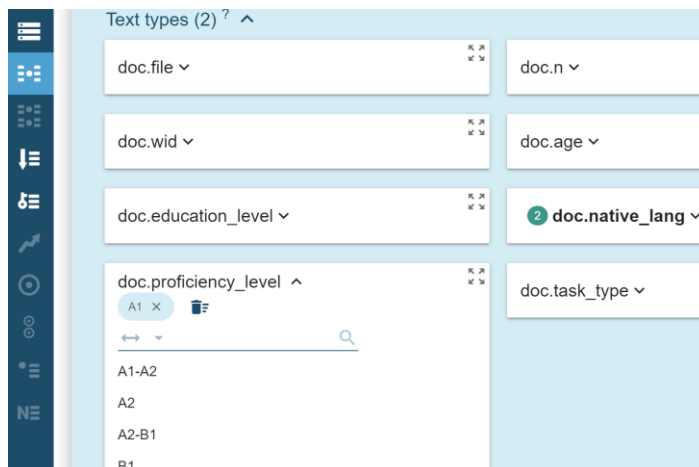


Figura 9

Pot fi făcute căutări avansate. În Fig. 10, am folosit opțiunea ADVANCED, apoi am selectat CQL și am aplicat formula de interogare [tagxpos="A.*"] pentru a găsi toate adjectivele din textele scrise de nativii bulgari și greci. Rezultatele sunt în Fig. 11.

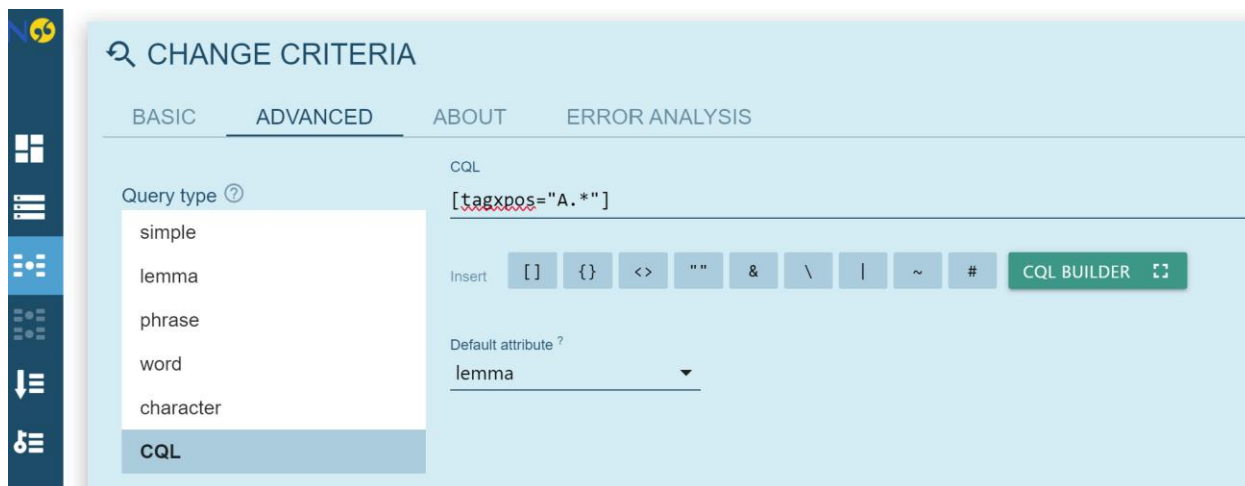


Figura 10.

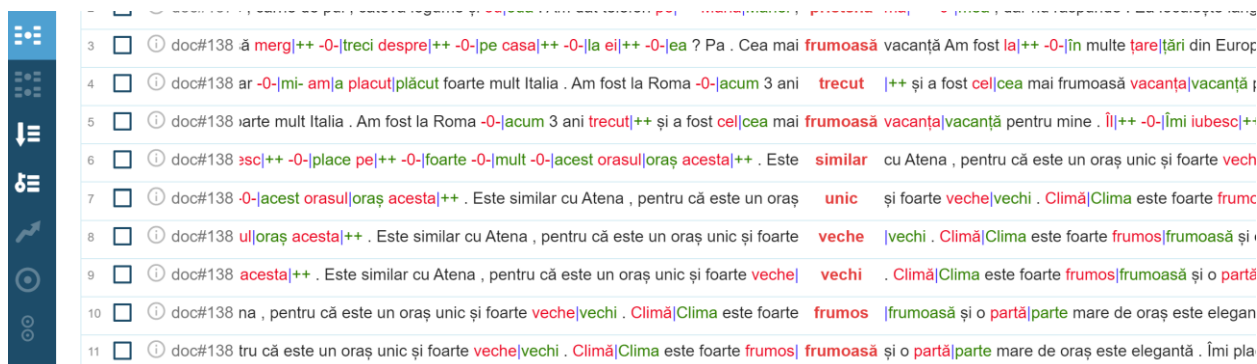


Figura 11

Exemplele pot fi salvate, bifând căsuțele din stânga (v. Fig. 12) și pot fi copiate apăsând pe simbolul de la capătul drept al rândului.

| | | | | | | | | |
|----|-------------------------------------|---------|------------------------------------|---|---------------------------|---------------------------|---------------------------|-------------------|
| 9 | <input checked="" type="checkbox"/> | doc#138 | acesta ++ | . Este similar cu Atena , pentru că este un oraș unic și foarte | veche | vechi | . Climă Clima este foarte | frumos fr |
| 10 | <input type="checkbox"/> | doc#138 | na , | pentru că este un oraș unic și foarte | veche vechi | . Climă Clima este foarte | frumos | frumoasă și o |
| 11 | <input checked="" type="checkbox"/> | doc#138 | tru că este un oraș unic și foarte | veche vechi | . Climă Clima este foarte | frumos | frumoasă | și o |
| 12 | <input type="checkbox"/> | doc#138 | i foarte | veche vechi | . Climă Clima este foarte | frumos frumoasă | și o | partă parte |
| 13 | <input type="checkbox"/> | doc#138 | i . | Climă Clima este foarte | frumos frumoasă | și o | partă parte | mare de oraș este |
| 14 | <input type="checkbox"/> | doc#138 | este elegantă . | Îmi place foarte mult pentru că are monumente care sunt foarte | unice | , că ca | Fontana di | Treci Trevi sau |
| 15 | <input type="checkbox"/> | doc#138 | foarte unice | , că ca | Fontana di | Treci | | |
| 16 | <input type="checkbox"/> | doc#138 | ntana di | Treci Trevi | sau | Colosseum | | |

Select on page: ALL NONE show only: SELECTED (2) NOT SELECTED COPY (2)

Figura 12

Figurile 13 și 14 ilustrează o căutare avansată cu lema „merge”, în texte scrise de studenți cu nivelul de competență A1, A1-A2 și A2 (doc.proficiency_level) scrise la examen (doc.task_type).

Figura 13

Figura 14

Descoperim informații detaliate despre student și despre text, dacă facem click pe abrevierea de tipul doc#138 din capătul stâng al fiecărui exemplu. Pentru doc#138, putem găsi informații de tipul celor din Fig. 15.

| | | |
|--------------------------|-----------------------------------|---|
| <input type="checkbox"/> | doc.native_lang | Greek |
| <input type="checkbox"/> | doc.no_of_words_s | 111 |
| <input type="checkbox"/> | doc.no_of_words_t | 113 |
| <input type="checkbox"/> | doc.particular_pragmatic_function | no |
| <input type="checkbox"/> | doc.proficiency_level | A2 |
| <input type="checkbox"/> | doc.ref_source_usage | Most probably yes |
| <input type="checkbox"/> | doc.sid | 75 |
| <input type="checkbox"/> | doc.task_mode | written |
| <input type="checkbox"/> | doc.task_requirements_ad_material | None |
| <input type="checkbox"/> | doc.task_type | work done in an exam |
| <input type="checkbox"/> | doc.text_function | narration |
| <input type="checkbox"/> | doc.topic | free time, entertainment, hobbies, travel |
| <input type="checkbox"/> | doc.wid | 8 |
| <input type="checkbox"/> | doc.writing_type | hand-written |

[SHOW ALL](#)

Figura 15

Pentru a avea acces la varianta originală a textului, facem click pe simbolul din capătul drept al fiecărui rând (sunt două tipuri de simboluri, pentru imagini făcute pdf, ca în primele două rânduri, și pentru fișiere audio, ca pentru următoarele două rânduri, v. Fig. 16):









| | | |
|------------------|---|---|
| severă | ? Răzgândește -te acum ! Fiica : Dar cât dorești să obțină o notă bună în ++ -0 |   |
| bună | în ++ -0 din inimă . De ce nu -l mângâi ca să se ușureze ? Mama : Și tu explici |   |
| moderni | este foarte stresantă . Dacă vrem să ne asigurăm sănătatea mintală , trebuie s |   |
| stresantă | . Dacă vrem să ne asigurăm sănătatea mintală , trebuie să căutăm ++ -0 acori |   |

Figura 16


În Fig. 17 sunt două exemple de variante originale ale textelor scrise de mână; acestea se pot salva în formatul pdf.

a) "Dacă aș câștiga la loto"

Dacă aș câștiga la loto aș merge ~~la~~ la Maldiv ca un vacanța cu prietenii mei. Trebuie să cumpăr biletele ~~pe~~ pe nave cel mai scump la nivelul început cu mai mult opțiune luxuroasă. Vreau să fac un duș cald lângă stele în avionul și după ce vreau să dorm în patul cel mai mare și confortabil toate noaptea. După aș fi ajuns la maldiv trebuie să lan un Vila atât de mare lângă plaje cu trei etaje și o piscină, cu un masă pentru tenis de masă și un teren de baschet. Vreau să fac un masaj la plajă și un trip cu avionul mică în jurul Maldivei. După această vacanță, vreau să merg la firma de Porsche și ~~eu~~ să cumpăr o mașină cel mai rapidă, acesta mașină este Porsche 911. Pentru facem mai mult de bani în viitorul, o să deschid o firmă internațională, această firmă ~~o~~ o să fie cel mai mare în România pentru hainele și ciocolată, în fiecare weekend, ~~pot~~ pot să lua prietenii mei la mal și pot să cumpere la mal ~~și~~. Și pentru mine o să deschid un spital cel mai mare și e ieftin pentru oameni care nu au bani pentru să vreau să fi un doctor. Mă place să câștig la loto foarte mult.

În vară trecută am mers la mare înConstanța cu prietenii mei, și am făcut ^{mult} acolo câț ~~și~~ sunt multe locuri de ~~facut~~ făcut, faci, ca poți să învâți și mergi într-o petrecere, și faci tan, iar și în vară trecută am mers și la munte și am urcat pe munte, a fost un experiență frumoasă, dar foarte obsit, am luat aproape 3 ~~ore~~ ore pentru ~~urcare~~ Dar am văzut câteva vedere frumoase și ~~și~~ am luat câteva poze, iar pentru mine cea mai Preferat destinația este mare, pentru ca ~~și~~ nu este tot atât de obsit ca la munte și acolo poți să întâlnești oameni, sper ca o să pot să merg în vară acesta.

Figura 17

2) Pentru al doilea tip de inițiere a căutării, se pornește de la opțiunea Text type analysis, apoi se alege un anumit tip de texte, de exemplu, textele în funcție de limba nativă (doc-nativ_lang), apoi se selectează o valoare, de exemplu texte scrise de nativi chinezi (v. Fig. 18). În rezultate (v. Fig. 19), se pot face alte căutări; o variantă este folosirea opțiunii de filtrare  (din linia de opțiuni din dreapta sus). În Fig. 19 s-a făcut o căutare simplă, după cuvântul „bun”.

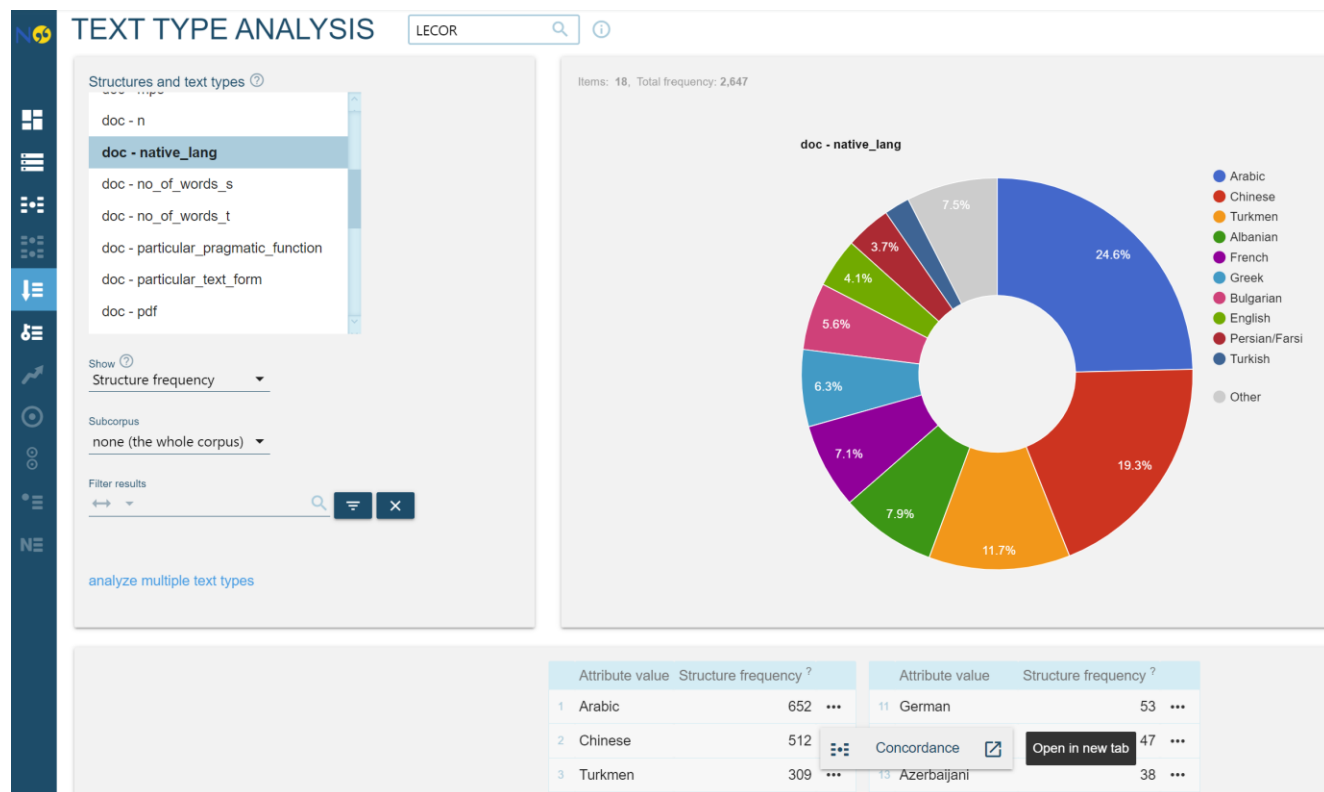


Figura 18

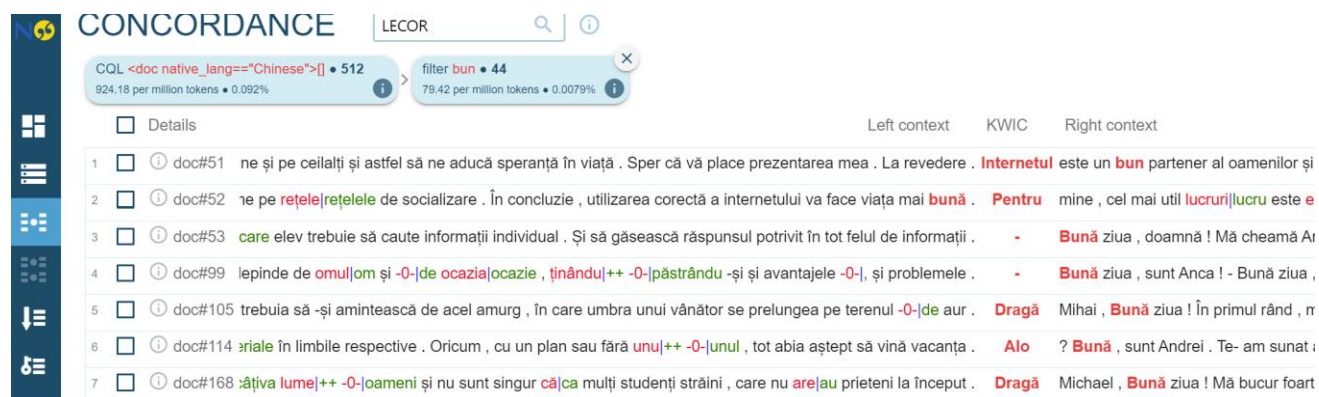


Figura 19

► În LECOR (primele lui trei variante prezentate) se pot face și **căutări ale erorilor**, restrângând după aceea tipurile de texte, în modurile exemplificate anterior.

În Fig. 20 este ilustrată o căutare avansată după elementele marcate ca greșite (WRONG) în tot corpusul, iar rezultatele sunt similare cu cele din Fig. 21.

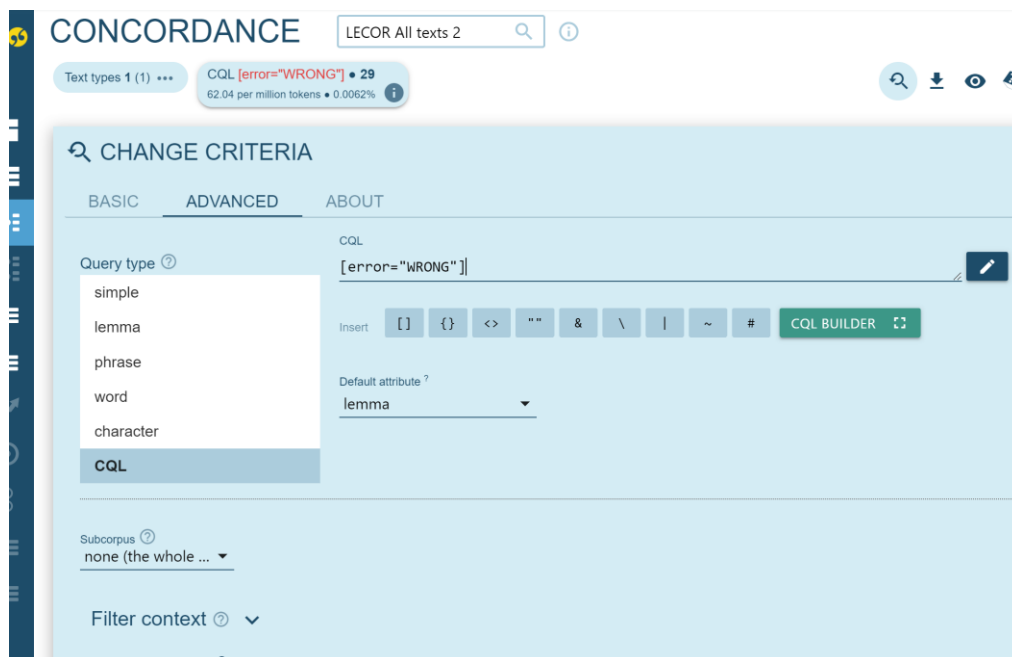


Figura 20

| Details | Left context | KWIC | Right context |
|---------------------------------------|---|--|---------------|
| 1 <input type="checkbox"/> ⓘ doc#145 | hoteluri , și multi lucruri pentru distracție . 12 iulie . | Astazi a fost sămbăta . Nu am mers la plaja , am preferăm | |
| 2 <input type="checkbox"/> ⓘ doc#145 | și multi lucruri pentru distracție . 12 iulie . Astazi a fost | sămbăta . Nu am mers la plaja , am preferăm să stăm puțin la t | |
| 3 <input type="checkbox"/> ⓘ doc#145 | ie . 12 iulie . Astazi a fost sămbăta . Nu am mers la | plaja , am preferăm să stăm puțin la hotel la piscina , pentru | |
| 4 <input type="checkbox"/> ⓘ doc#145 | Astazi a fost sămbăta . Nu am mers la plaja , am | preferăm să stăm puțin la hotel la piscina , pentru că am vrut să merge | |
| 5 <input type="checkbox"/> ⓘ doc#145 | am mers la plaja , am preferăm să stăm puțin la hotel la | piscina , pentru că am vrut să mergem la orașul unsulei seara . La | |
| 6 <input type="checkbox"/> ⓘ doc#145 | n la hotel la piscina , pentru că am vrut să mergem la orașul | unsulei seara . La oră 19:00 am mers la orașul cu 2 mașine | |
| 7 <input type="checkbox"/> ⓘ doc#145 | , pentru că am vrut să mergem la orașul unsulei seara . La | oră 19:00 am mers la orașul cu 2 mașine -0- care -0- am | |
| 8 <input type="checkbox"/> ⓘ doc#145 | mergem la orașul unsulei seara . La oră 19:00 am mers la | orașul cu 2 mașine -0- care -0- am inchiriat . Am facut | |
| 9 <input type="checkbox"/> ⓘ doc#145 | unsulei seara . La oră 19:00 am mers la orașul cu 2 | mașine -0- care -0- am inchiriat . Am facut o plimbare și a | |
| 10 <input type="checkbox"/> ⓘ doc#145 | 19:00 am mers la orașul cu 2 mașine -0- care -0- am | inchiriat . Am facut o plimbare și a fost foarte bine . La ora 22:00 : | |

Figura 21

În Fig. 22, sunt ilustrate rezultatele unei căutări avansate similare celei anterioare, dar vizând erorile de omisiune (etichetate cu simbolul -0-). Rezultatele includ elemente care, în mod greșit, sunt omise de student sau care au o ordine greșită. De exemplu, în primele două rânduri din figură (Fig. 22), sunt omise morfemul de conjunctiv „să”, respectiv prepoziția funcțională „pe”, iar în al treilea rând, poziționarea adjectivului este greșită („ziua ultimă de vacanță”).

The screenshot shows a list of search results for the query 'doc#145'. Each line contains a snippet of text with various error markers (e.g., -0-, OK, WRONG, RED+/-) and icons for document details. The text snippets are:

- 10: doc#145 rsoane , la -0- toate -0- poate să -0- -0- -0- -0- -0- aibă 12 persoane . Nu a fost foarte scump și am ir
- 11: doc#145 mp și am inchiriat iahtul pentru o zi . Am mers la alte plaje , -0- care nu putem să mergem cu mașina . A fost minunat . 15 iulie .
- 12: doc#145 nu putem să mergem cu mașina . A fost minunat . 15 iulie . Este -0- ziua ultimă de vacanță . Am fost obosiți -0- dar și mulțumiți p
- 13: doc#145 5 iulie . Este -0- ziua ultimă de vacanță . Am fost obosiți -0- dar și mulțumiți pentru -0- vacanța . Am mers puțin la oraș și
- 14: doc#145 ja ultimă de vacanță . Am fost obosiți -0- dar și mulțumiți pentru -0- vacanța . Am mers puțin la oraș și după -0- am mâncat .
- 15: doc#145 și mulțumiți pentru -0- vacanța . Am mers puțin la oraș și după -0- am mâncat . M- am întors la hotelul și -0- Pregatim :
- 16: doc#145 raș și după -0- am mâncat . M- am întors la hotelul și -0- Pregatim să mergem la Atena . Locuița de vis . Casa ide

Figura 22

În Fig. 23, este ilustrată o căutare avansată, cu <err/> [tagxpos="Ncf.*"], care vizează substantivele comune de genul feminin (Ncf.*) greșite. Aceste rezultate pot fi ulterior limitate la tipurile de texte care interesează.

The screenshot shows the CONCORDANCE interface with the search query 'CQL <err/> [tagxpos="Ncf.*'] • 15,495'. The results are displayed in a table with columns for document ID, left context, KWIC, and right context. The results are:

- 1: doc#0 în imagina | imaginea aceasta sunt două femei . Po
- 2: doc#1 îngă și albă și nul++ -0-|un jachet maro foarte frumos . Este o zi bună să facă o plimbare ! În imaginea | imagine se vede o femeie extrem de fe
- 3: doc#1 și soție , În ziua lor de căsătorie . Gândindu -se la un loc mai special , unde să facă pozele din nunta | nuntă , au decis să fie creativi și și
- 4: doc#2 pentru prima dată . Ieri am avut o zi foarte grea . M- am trezit , ca de obicei , laora|++ -0-|la -0- | ora 6 . Am făcut un duș , m- ar
- 5: doc#2 trezit , ca de obicei , laora|++ -0-|la -0-|ora 6 . Am făcut un duș , m- am spălat pe dinți și pe fata | față . Apoi m- am pieptănat . Arr
- 6: doc#2 dar espressorul ș- a stricat . Încă adormită , m- am îmbrăcat și m- am încălțat . M- am dus la facultatea | facultate . M- am grăbit și am fugit duș
- 7: doc#2 cultate . M- am grăbit și am fugit după autobuz . Am urcat în autobuzul 365 . Apoi am ajuns la facultatea | facultate și m- am întâlnit cu prietena m

Figura 23

► **Vizualizarea** rezultatelor în LECOR se poate face în variantele generale disponibile pentru interfața (no)Sketch engine. Fig. 24 și 25 ilustrează un tip de vizualizare cu mai puține informații, iar Fig. 26 și 27, cu mai multe informații.

The screenshot shows the 'VIEW OPTIONS' panel in the CONCORDANCE interface. It includes sections for 'Show attributes?' and 'Show structures?'. Under 'Show attributes?', there are checkboxes for 'words (1)', 'lemmas (2)', 'tagupos', 'error', 'eseq', 'stypc (3)', and 'words (lowercase)'. Under 'Show structures?', there are checkboxes for 'err' and 'corr'. There are also options for 'For each token' and 'For KWIC only', and a 'Show as tooltips' checkbox. At the bottom, there are several toggle switches for 'Show more context?', 'Show line selection', 'Show line copy icon', 'Show line numbers?', and 'Use glue?'.

Figura 24

CONCORDANCE LECOR

CQL <err> [tagxpos="Ncf.*"] • 15,495
20,536.78 per million tokens • 2.8%

Details Left context KWIC Right context

| | | | | |
|---|-------|--|------------------------|--|
| 1 | doc#0 | în | imagina imaginea | aceasta sunt două femei . Pot să văd că ele sunt în |
| 2 | doc#1 | hie lungă și albă și nu++ -0- un jachet maro foarte frumos . Este o zi bună să facă o plimbare ! În | imaginea imagine | se vede o femeie extrem de fericită , care poartă o roc |
| 3 | doc#1 | soț și soție , în ziua lor de căsătorie . Gândindu -se la un loc mai special , unde să facă pozele din | nunta nuntă | au decis să fie creativi și și- au rezervat un stadion |
| 4 | doc#2 | âlnit pentru prima dată . Ieri am avut o zi foarte grea . M- am trezit , ca de obicei , laora++ -0- la | -0- ora | 6 . Am făcut un duș , m- am spălat pe dinți și pe fata fat |
| 5 | doc#2 | am trezit , ca de obicei , laora++ -0- la -0- ora 6 . Am făcut un duș , m- am spălat pe dinți și pe | fata față | Apoi m- am pieptănat . Am vrut să face fac o cafea , |
| 6 | doc#2 | dar espressoorul s- a stricat . Încă adormită , m- am îmbrăcat și m- am încălțat . M- am dus la | facultatea facultate | M- am grăbit și am fugit după autobuz . Am urcat în a |
| 7 | doc#2 | ifacultate . M- am grăbit și am fugit după autobuz . Am urcat în autobuzul 365 . Apoi am ajuns la | facultatea facultate | și m- am întâlnit cu prietena mea . M- am bucurat , dar |
| 8 | doc#2 | mea . M- am bucurat , dar din nefericire am avut o notă proastă la test și m- am supărat . După | curturile cursuri | am plecat acasa acasă , m- am plimbat în parcul -0- de |

Figura 25

CONCORDANCE LECOR

CQL <err> [tagxpos="Ncf.*"] • 15,495
20,536.78 per million tokens • 2.8%

VIEW OPTIONS

Show attributes ? (5)
type to search

- words (1)
- lemmas (2)
- tagupos (4)
- tagupos (5)
- error
- eseq
- etyp (3)
- words (lowercase)

Show structures ? (2)
type to search

- err
- corr
- p
- doc

Show more context ?

Show line selection

Show line copy icon

Show line numbers ?

Use glue ?

LINE DETAILS

SAVE

Figura 26

CONCORDANCE LECOR

CQL <err> [tagxpos="Ncf.*"] • 15,495
20,536.78 per million tokens • 2.8%

Details Left context KWIC Right context

| | | | | |
|----|-------|---------------------------------|------------------------|-------------------------|
| 1 | doc#0 | în | imagina imaginea | aceasta sunt două femei |
| 2 | doc#1 | bună să facă o plimbare ! În | imaginea imagine | se vede o femeie |
| 3 | doc#1 | special unde să facă pozele din | nunta nuntă | au decis să |
| 4 | doc#2 | ca de obicei laora++ -0- la | -0- ora | 6 Am făcut |
| 5 | doc#2 | am spălat pe dinți și pe | fata față | Apoi m- am |
| 6 | doc#2 | încălțat M- am dus la | facultatea facultate | M- am grăbit |
| 7 | doc#2 | juzul 365 Apoi am ajuns la | facultatea facultate | și m- am întâlnit |
| 8 | doc#2 | și m- am supărat După | curturile cursuri | am plecat acasa acasă |
| 9 | doc#2 | it în parcul -0- de lângă | -0- casă | Am revenit acasa |
| 10 | doc#3 | am cultat culcat | Memorii ++ | -0- Amintiri |

Figura 27

În LECOR, pot fi vizualizate și erori speciale (lexico-gramaticală), cu opțiunea etype (v. Fig. 28).

The screenshot shows the LECOR concordance interface. At the top, there's a search bar with 'LECOR' and a magnifying glass icon. Below it, there's a filter for 'Text types 1 (1) ... simple . • 22' and '29.16 per million tokens • 0.0029%'. The main area displays 15 search results for the word 'insulă'. Each result shows a document ID (doc#215), the word 'insulă' with its error tags (FS, GAE, LSV), and the surrounding text. The error tags are color-coded: FS (green), GAE (red), and LSV (blue). The interface also has a sidebar on the left with various navigation icons and a 'Details' button at the top left.

Figura 28

Lista de erori speciale lexico-semantice are 62 de elemente (v. Fig.29). S-a pornit de la o listă similară, construită pentru engleza ca limbă nenativă (*Louvain Error Tagging Manual*, version 2.0, Granger, Swallow & Thewissen, 2022²). Lista a fost adaptată specificului lexico-gramatical al limbii române.

| domain / level | | Error tags | Error tag name |
|--|----|---------------|---|
| 1. Formal errors (2) 1-2 | 1 | FDM | FDM: Form, derivational morphology |
| | 2 | FS | FS: Form, spelling |
| 2. Grammatical errors (33) 3-35 | 3 | GAE | GAE: Grammar, article, enclitic |
| | 4 | GAI | GAI: Grammar, article, indefinite |
| | 5 | GAO | GAO: Grammar, article, possessive |
| | 6 | GAD | GAD: Grammar, article, demonstrative |
| | 7 | GDD | GDD: Grammar, determiner, demonstrative |
| | 8 | GDDO | GDDO: Grammar, determiner, demonstrative, order |
| | 9 | GDO | GDO: Grammar, determiner, possessive |
| | 10 | GDOO | GDOO: Grammar, determiner, possessive, order |
| | 11 | GDI | GDI: Grammar, determiner, indefinite |
| | 12 | GDNEG | GDNEG: Grammar, determiner, negative |
| | 13 | GDNUM | GDNUM: Grammar, determiner, numeral |
| | 14 | GDT | GDT: Grammar, determiner, other |
| | 15 | GDT | GDT: Grammar, determiner, other |
| | 16 | GADJCS | GADJCS: Grammar, adjective, comparative/superlative |

² <https://uclouvain.be/en/research-institutes/ilc/cecl/cecl-papers.html>

| | | | |
|---|----|---------------|--|
| | 17 | GADM | GADM: Grammar, adjective, inflectional morphology (number, gender, case) |
| | 18 | GADJO | GADJO: Grammar, adjective, order |
| | 19 | GADVCS | GADJCS: Grammar, adverb, comparative/superlative |
| | 20 | GADVO | GADVO: Grammar, adverb, order |
| | 21 | GNM | GNM: Grammar, noun, inflectional morphology (number, gender, case) |
| | 22 | GPD | GPD: Grammar, pronoun, demonstrative |
| | 23 | GPI | GPI: Grammar, pronoun, indefinite |
| | 24 | GPR | GPR: Grammar, pronoun, relative/ interrogative |
| | 25 | GPI | GPI: Grammar, pronoun, indefinite |
| | 26 | GPP | GPP: Grammar, pronoun, personal |
| | 27 | GPF | GPF: Grammar, pronoun, reflexive/reciprocal |
| | 28 | GPC | GPC: Grammar, pronoun, clitic |
| | 29 | GPU | GPU: Grammar, pronoun, unclear reference |
| | 30 | GVAUX | GVAUX: Grammar, verb, auxiliary |
| | 31 | GVM | GVM: Grammar, verb, morphology |
| | 32 | GVMT | GVMT: Grammar, verb, mood, tense |
| | 33 | GVV | GVV: Grammar, verb, voice |
| | 34 | GWC | GWC: Grammar, word class |
| | 35 | GDOM | GDOM: Grammar, direct object marker |
| 3. Lexical errors (10) 36-45 | 36 | LCC | LCC: Lexis, conjunction, coordination |
| | 37 | LCLC | LCLC: Lexis, connector, logical, complex |
| | 38 | LCLS | LCLS: Lexis, connector, logical, single |
| | 39 | LCS | LCS: Lexis, conjunction, subordination |
| | 40 | LP | LP: Lexical phrase errors |
| | 41 | LSADJ | LSADJ: Lexical single adjective |
| | 42 | LSADV | LSADV: Lexical single adverb |
| | 43 | LSN | LSN: Lexical single noun |
| | 44 | LSPR | LSPR: Lexical single preposition |
| | 45 | LSV | LSV: Lexical single verb |
| 4. Punctuation errors (4) 46-49 | 46 | QC | QC: Punctuation, confusion |
| | 47 | QL | QL: Punctuation, lexical |
| | 48 | QM | QM: Punctuation, missing |
| | 49 | QR | QR: Punctuation, redundant |
| 5. Word missing/redundant/order errors (3) 50-52 | 50 | WM | WM: Word missing |
| | 51 | WO | WO: Word order |
| | 52 | WR | WR: Word redundant |
| 6. Lexico-grammatical errors (9) 53-61 | 53 | XADJCO | XADJCO: Lexico-grammar, adjective, complementation |
| | 54 | XADJPR | XADJPR: Lexico-grammar, adjective, preposition |
| | 55 | XADVCO | XADVCO: Lexico-grammar, adverb, complementation |
| | 56 | XADVPR | XADVPR: Lexico-grammar, adverb, preposition |
| | 57 | XNCO | XNCO: Lexico-grammar, noun, complementation |
| | 58 | XNPR | XNPR: Lexico-grammar, noun, preposition |
| | 59 | XPRCO | XPRCO: Lexico-grammar, preposition, complementation |
| | 60 | XVCO | XVCO: Lexico-grammar, verb, complementation |
| | 61 | XVPR | XVPR: Lexico-grammar, verb, preposition |
| 7. Infelicities (1) 62 | 62 | Z | Z: Infelicities |

Figura 29