

UTILIZAREA INDICELUI CTTR ÎN ANALIZA DIVERSITĂȚII LEXICALE A INTERLIMBII STUDENȚILOR CARE ÎNVAȚĂ RLS

Mariana-Diana CÂȘLARU*

Elena-Mihaela ANDREI**

Universitatea „Alexandru Ioan Cuza” din Iași

*dianacaslaru@yahoo.fr

**myhaela_andrey@yahoo.com

The Use of the CTTR Index in Lexical Diversity Analysis of RFL Students' Interlanguage

DOI: 10.35923/AUTFil.59.02

The transitory nature of interlanguage, seen as a temporary linguistic system of a foreign language learner, requires the use of methodologies that go beyond the narrow framework of compliance with the norm. In this article, we aim to present a methodology for analyzing the lexical diversity of interlanguage, using the CTTR index. We will also present the results of a pre-test that applies this approach to the language production of international RFL students.

Keywords: *lexical diversity; interlanguage; type-token ratio; Romanian as a Foreign Language; target language.*

Cadru teoretic despre interlimbă și diversitatea lexicală în interlimbă

Ipoteza interlimbii reprezintă o încercare – dovedită a fi una fecundă – de a depăși majoritatea abordărilor anterioare privitoare la procesul de învățare a unei limbi străine. Original în ipotezele lui, conceptul de interlimbă propune o nouă perspectivă de a înțelege fenomenul învățării unei alte limbi. Conform acestei noi perspective, se consideră că, atunci când un subiect învață o altă limbă, acesta dezvoltă un alt sistem lingvistic, diferit atât de limba maternă, cât și de limba țintă, dar care conține elemente proprii celor

două. De aici, a fost necesar un singur pas până la a spune că limba subiectului sau a elevului se comportă ca un sistem intermediar, sistem aproximativ, gramatică/ competență tranzitorie, dialect idiosincrazic sau interlimbă¹ (termenul cel mai uzitat de către cercetători).

Pit Corder (1980a, 1980b și 1980c), ale cărui lucrări dedicate interlimbii au constituit o trambulină pentru lansarea acestei noi abordări, este de părere că limba elevului nu este un dialect social, ci un dialect individual.

Mai mult decât atât, în interlimbă, eroarea este o „binecuvântare” (Kiyit-sioglu-Vlachou 2001), ea demonstrând existența activității cognitive de a verifica ipoteze asupra funcționării limbii studiate. Această nouă abordare are ca obiectiv principal descrierea gramaticii interioare a subiectului (Besse și Porquier 1991), descriere care trebuie să se facă, într-o primă etapă cel puțin, în propriii săi termeni, fără a face apel la norma limbii țintă.

„Son principal objectif est en effet de décrire les grammairiales intériorisées à travers les activités langagières qui les manifestent, pour en caractériser les spécificités, les propriétés et les modalités de leur développement.” (Besse et Porquier 1991: 216).

Dacă o analiză corectă a interlimbii vizează strict sistemele provizorii, fără a le raporta la limba țintă, atunci una dintre dimensiunile pe care ar trebui să le analizăm este diversitatea lexicală, deoarece aceasta va indica nu doar cât de divers este vocabularul subiectului care învață o limbă străină, ci și puterea sa de a formula ipoteze despre sistemul lingvistic țintă. Atunci când acel „savoir” interlingual este insuficient, adică subiectul nu cunoaște un cuvânt, el va căuta să îl înlocuiască cu un termen care se va constitui fie într-un calc/substituție, fie într-un transfer pozitiv. Indiferent de rezultat, transfer pozitiv sau negativ, acesta este perfect acceptabil, întrucât interlimba este permeabilă și mereu în schimbare. Așadar, în calculul diversității lexicale a interlimbii, vor fi incluse și acele forme care nu sunt conforme cu norma limbii țintă, ele fiind semn al activității cognitive a subiectului și nicidecum indici ai eșecului.

¹ Acest termen a fost propus de lingvistul american Larry Selinker: „This set of utterances for most learners of a second language is not identical to the hypothesized corresponding set of utterances which would have been produced by 40 native speaker of the TL had he attempted to express the same meaning as the learner. since we can observe that these two sets of utterances are not identical, [...], one would be completely justified in hypothesizing, perhaps even compelled to hypothesize, the existence of a separate linguistic system based on the observable output which results from a learner's attempted production of a TL norm. This linguistic system we will call *interlanguage*”. (Selinker, 1972: 214)

În acest sens, Pit Corder afirma că nu putem numi ca „deviante” sau „incorecte” producțiile unui copil care este pe cale să dobândească limba sa maternă, întrucât el nu este încă locutorul unui dialect social. În mod analog, nu putem numi „deviante” sau „eronate” producțiile idiosincrazice ale unui subiect care învață o limbă străină, întrucât aceste etichetări implică / sugerează nerespectarea regulilor limbii țintă, or subiectul produce structuri idiosincrazice tocmai pentru că nu cunoaște încă aceste reguli sau nu le-a asimilat/fixat/interiorizat încă (1980b).

Metodologie de analiză a diversității lexicale

Cea mai populară formulă pentru calcularea diversității lexicale a unui text este TTR (type/token ratio). Este bine cunoscut faptul că TTR este influențat negativ de lungimea textului analizat, în sensul că pe măsură ce crește dimensiunea textului, scade indicele diversității lexicale, întrucât cuvintele tind să se repete odată cu creșterea lungimii textului. După ce s-au adus numeroase critici utilizării indicelui TTR (Wolf Quintero *et alii* 1998), în 1964 s-a propus o nouă formulă de calcul, numită CTTR (Corrected Type Token Ratio) de către Carroll (1964: 54), care este în directă legătură cu lungimea textului. Această formulă este agreată și de cercetătorul contemporan Mathias Schulze (articol inedit), care conduce numeroase proiecte de cercetare a interlimbii în diferite centre universitare din Canada.

Indicele CTTR² calculează numărul de tipuri de cuvinte împărțit la rădăcina pătrată din dublul numărului de cuvinte. Ocurențele identice ale unui cuvânt sunt numărate ca fiind un singur tip (casă, casă = 1 tip, iar casă, case = 2 tipuri).

$$CTTR = T \div \sqrt{2 \times W}$$

unde T = tipuri de cuvinte, iar W = numărul total de cuvinte

Pentru a calcula corect numărul de tipuri de cuvinte, propunem utilizarea programului Word List Expert, dezvoltat de Mechanicwords. Precizăm că este necesară eliminarea cratimelor din corpusurile studenților, din structuri precum *l-am*, *m-ai*, *te-au*, deoarece am constatat că ele sunt numărate ca reprezentând un singur cuvânt.

² O descriere mult mai detaliată a acestei formule poate fi găsită în Mariana Diana Câșlaru, *L'interlangue des apprenants roumains de FLE au carrefour des langues romanes (études de cas sur des apprenants roumains étudiant aussi l'italien et l'espagnol)*, Presa universitară clujeană, 2016.

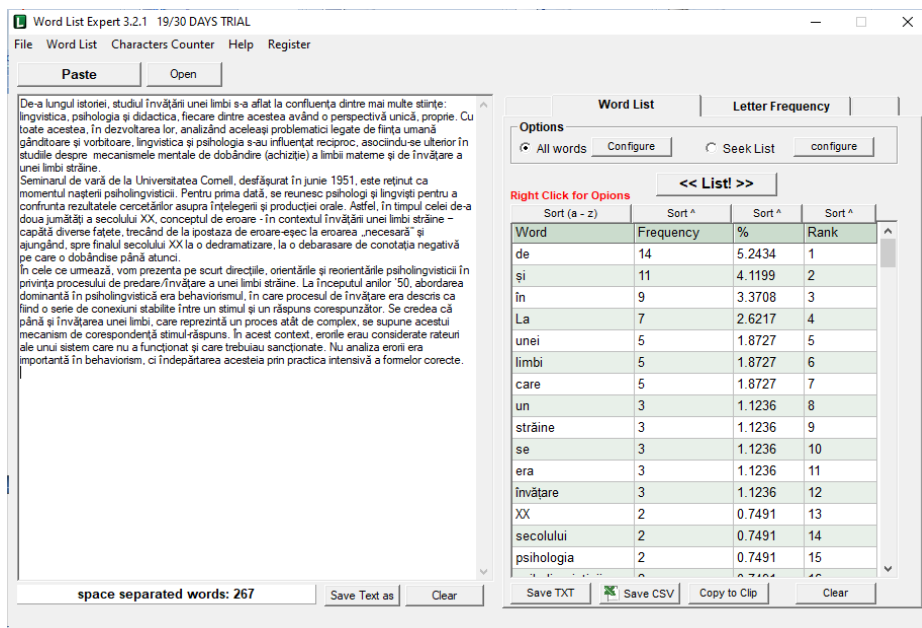


Figura 1. Exemplu de numărare a tipurilor de cuvinte

Întrucât interlimba este un continuum de sisteme, această procedură de calcul al indicelui diversității lexicale trebuie să fie utilizată asupra mai multor producții scrise /orale din cadrul unei interlimbi, pentru a putea obține o imagine în diacronie a interlimbii în cauză.

Eșantioane de analiză a diversității lexicale

În cele ce urmează, vom prezenta rezultatele etapei de pretestare a acestei metodologii de analiză a diversității lexicale a interlimbii studenților care învață limba română ca limbă străină. Am avut ca subiecți un număr de zece studenți internaționali, înscriși la Programul de An pregătitor, din cadrul Facultății de Litere a Universității „Alexandru Ioan Cuza” din Iași, cu un nivel de limbă română A2 – B1. Am efectuat trei anchete în decurs de șase luni, pe parcursul anului universitar 2020-2021. Vom prezenta rezultatele a trei dintre aceștia, proveniți din Togo, Sudan și Nigeria.

În urma utilizării softului Word List Expert și după calculul indicelui CTTR, am obținut datele din următorul tabel, pe care le-am reprezentat în graficele de mai jos, pentru fiecare subiect în parte.

	Subiectul 1			Subiectul 2			Subiectul 3		
Textul	Cuvinte	Tipuri	CTTR	Cuvinte	Tipuri	CTTR	Cuvinte	Tipuri	CTTR
1	199	131	6.56	163	106	5.87	157	102	5.75
2	375	183	6.68	411	188	6.55	273	147	6.29
3	357	190	7.11	347	167	6.33	315	137	5.45

Tabel 1. Calcule Subiectul 1, Subiectul 2 și Subiectul 3

Subiectul 1 – interpretarea rezultatelor

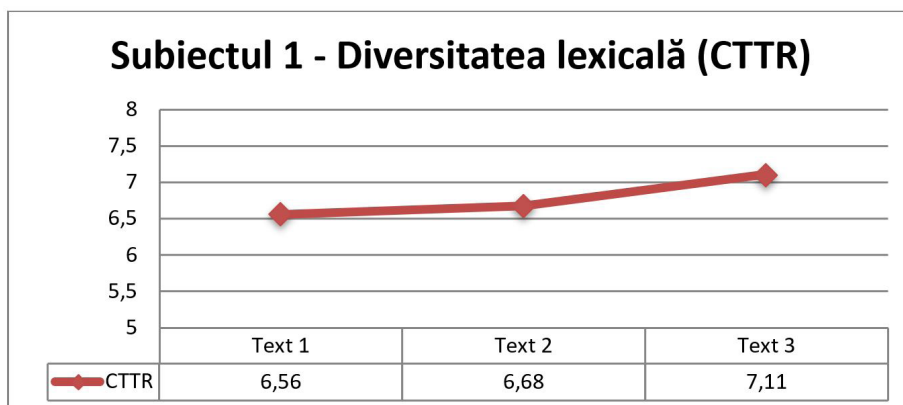


Figura 2. Diversitatea lexicală – subiectul 1

În cazul primului subiect, observăm o creștere continuă a diversității lexicale de-a lungul celor trei anchete, ceea ce înseamnă că subiectul își dezvoltă vocabularul din ce în ce mai mult pe parcursul celor șase luni de studiu. Este de remarcat faptul că între momentul 2 și momentul 3, diversitatea lexicală a interlimbii subiectului 1 înregistrează o creștere mai abruptă decât între momentele 1 și 2. Această variație poate fi datorată unei reale îmbogățiri a vocabularului subiectului nostru sau poate fi explicată printr-o mai mare siguranță de sine, o mai mare dezinvoltură a studentului în momentul producerii textului, ceea ce duce implicit la o mai largă plajă de ipoteze formulate asupra sistemului limbii țintă, ipoteze care se pot concretiza în transferuri negative sau pozitive și care sunt înregistrate ca îmbogățiri la nivel lexical. Chiar dacă o anumită formă lexicală propusă de subiect nu este 100% conformă cu norma limbii țintă, ea intră în calculul

diversității lexicale, deoarece, așa cum am stabilit de la început, analiza de față privește interlimba în afara raportului său cu sistemul limbii țintă. De altfel, dacă privim și din perspectivă didactică, preferăm ca un student să propună o structură lexicală chiar și imperfectă, decât să lase goluri în producția sa.

Subiectul 2 – interpretarea rezultatelor

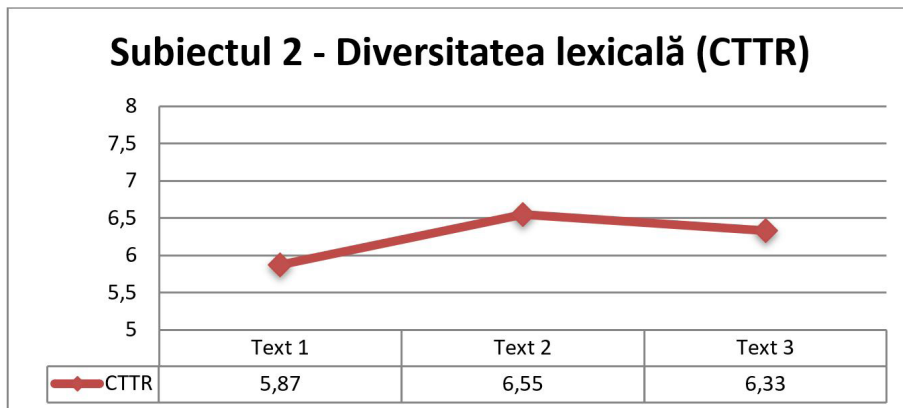


Figura 3. *Diversitatea lexicală – subiectul 2*

La cel de-al doilea subiect, constatăm o creștere abruptă a diversității lexicale între primele două momente ale anchetelor și o scădere mai lentă între ultimele două momente. Considerăm că subiectul 2 își diversifică lexicul în mod incontestabil de la momentul 1 la momentul 2, însă, în ceea ce privește scăderea indicelui diversității spre finalul anchetelor, nu am crede că este vorba de o involuție în interlimbă, ci mai degrabă de o alegere de moment. Subiectul 2 a încheiat Anul Pregător cu un nivel de limbă B1 spre B2, cadrele didactice constatând o dezvoltare remarcabilă a interlimbii acestuia în ansamblul ei.

Scăderile care nu sunt extreme, asemănătoare acestora de mai sus, ar putea avea două cauze. Prima este legată de starea fizică sau psihică a subiectului în momentul anchetei, stare care, după cum bine cunoaște un cadru didactic, poate influența performanța de moment a subiectului. A doua cauză este în strânsă legătură cu conformitatea interlimbii în raport cu norma limbii țintă. Unii studenți devin mai atenți la conformitatea cu norma, se concentrează să scrie corect, să utilizeze numai forme de care ei sunt foarte siguri, ceea ce conduce aproape inevitabil la scăderea diversității lexicale și sintactice. Pentru a stabili cu exactitate care dintre cele două cauze a determinat scăderea

indicelui diversității lexicale în contextul de față, ar trebui să continuăm analiza corpusului prin calcularea indicelui de conformitate cu norma limbii țintă.

Subiectul 3 – interpretarea rezultatelor

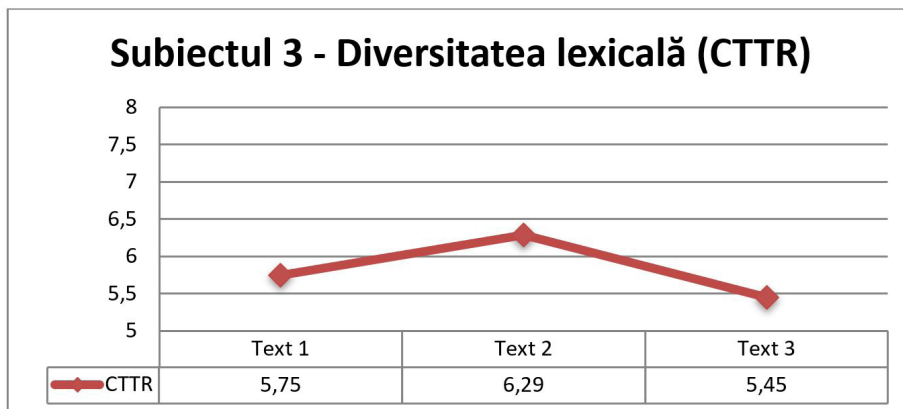


Figura 4. Diversitate lexicală - subiectul 3

În cazul celui de-al treilea subiect, traseul indicelui diversității lexicale urmează o curbă ascendentă între primele două momente ale studiului și una descendentă între ultimele două. Am încercat să căutăm cauzele scăderii indicelui în cea de-a treia anchetă. Privind textele 2 și 3, produse de către subiectul 3, considerăm că este posibil ca subiectele alese pentru a fi dezvoltate să fi influențat scăderea diversității lexicale întrucât în textul 2 subiectul relatează o serie de întâmplări diferite, petrecute în spații diferite, pe când în textul 3, subiectul povestește gradual o singură întâmplare, făcând mereu referire la aceleași personaje și, prin urmare, reluând mereu aceleași cuvinte.

Concluzii

Rezultatele pretestărilor ne indică faptul că acest indice CTTR poate fi utilizat, cu obținerea de rezultate fiabile, pe producțiile scrise ale studenților internaționali care învață RLS. Analiza diversității lexicale este necesară în cazul interlimbii, întrucât evoluția ei diferă de modul în care variază indicele de conformitate cu norma (indice luat cel mai des în considerare atunci când este evaluată o interlimbă). Or, noile metode de predare-învățare, centrate pe comunicare /acțiune, pun accentul pe renunțarea la taxarea erorii și pe evaluarea producțiilor studenților, fără a sublinia ceea ce nu este conform cu limba țintă, cel puțin într-o primă etapă.

Așadar, pentru o mai bună înțelegere a modului în care se dezvoltă interlimba, analiza producțiilor studenților poate și trebuie să fie extinsă prin corelarea indicelui conformității cu norma cu indicele diversității lexicale. Anticipăm că, făcând această corelare, vom constata (din experiența de cadre didactice cunoaștem acest lucru) că sunt momente în care un indice crește concomitent cu descreșterea celuiilalt. Aceste ușoare regrese fac parte din procesul de învățare, ele intrând deseori într-un mecanism compensatoriu. De exemplu, cu cât subiectul progresează în învățarea unui lexic variat, cu atât atenția lui poate scădea în privința corectitudinii gramaticale. În economia interlimbii, cu excepția cazurilor de fosilizare, orice scădere a unui indice poate ascunde un progres, întrucât subiectul face un pas înapoi pentru a se concentra pe o altă componentă.

Încheiem aceasta propunere metodologică, care s-a dovedit fiabilă, prin lansarea unei noi provocări nouă însene: aceea de a continua studiul interlimbii studenților care învață RLS, prin corelarea indicelui CTTR cu indicele conformității cu norma – ceea ce presupune elaborarea unui sistem de etichetare a erorilor, urmată de o etichetare manuală a acestora și apoi de calcularea asistată de calculator a indicelui conformității.

REFERINȚE BIBLIOGRAFICE:

- BESSE, Henry, PORQUIER, Rémy 1991: *Grammaires et didactique des langues*, Paris, Hatier-Didier.
- CARROLL, John 1964: *Language and Thought*, Englewood Cliffs, Prentice-Hall.
- CÂȘLARU, Mariana-Diana 2016: *L'interlangue des apprenants roumains de FLE au carrefour des langues romanes (études de cas sur des apprenants roumains étudiant aussi l'italien et l'espagnol)*, Presa universitară clujeană.
- CORDER, Pit 1980a: *Que signifient les erreurs des apprenants ?*, „Langages 57. Apprentissage et connaissance d'une langue étrangère”, pp. 9-15.
- CORDER, Pit 1980b: *Dialecte idiosyncrasique et l'analyse d'erreurs*, „Langages 57. Apprentissage et connaissance d'une langue étrangère”, pp. 17-28.
- CORDER, Pit 1980c: *La sollicitation de données d'interlangue*, „Langages 57. Apprentissage et connaissance d'une langue étrangère”, pp. 29-38.
- KIYITSIOGLU-VLACHOU, Catherine 2001: *Les bienfaits de l'erreur*, „Le français dans le monde”, 315, pp. 30-31.

SCHULZE, Mathias, VERSPOOR, Marjolijn, WOOD, Peter, POKORNY, Bjanka
(articol inedit), *Towards automatic proficiency scoring in L2 writing: Balanced
complexity*.

SELINKER, Larry 1972: *Interlanguage*, in „International review of applied lingu-
istics in language teaching” 10/3, pp. 209-231.

WOLF-QUINTERO, Kate, INAGAKI, Shunji, HAE-YOUNG, Kim 1998: *Second
language development in writing: measures of fluency, accuracy & complexity*,
University of Hawai‘i.