PARTICULARITĂȚI STILISTICE ȘI PRAGMATICE ALE TEXTELOR SOCIAL-MEDIA

CĂTĂLINA MĂRĂNDUC
Institutul de lingvistică „Iorgu Iordan – Al. Rosetti" al Academiei Române, București
Universitatea „Al. I. Cuza", Iași
CENEL AUGUSTO PEREZ
Universitatea „Al. I. Cuza", Iași

1. Introducere

1.1. *Prezentarea corpusului*

Autorii au prezentat în mai multe comunicări (Mărănduc și Perez 2015, Perez 2015) un corpus de tip treebank pe care l-au creat în cadrul colectivului de Procesare a Limbajului Natural (NLP) al Facultății de Informatică de la Iași. Corpusul este format acum din peste 11.000 de fraze sau propoziții, în format xml. Ele sunt obținute prin procesare automată a textului, apoi validate (corectate) manual.

În Figura 1 prezentăm un astfel de arbore. Numărul de identificare al fiecărui cuvânt sau semn de punctuație este urmat de forma ocurentă în text, de lemma, adică forma de bază din dicționar, apoi de analiza morfologică (postag = part of speech tag), în care fiecare literă codifică o informație. Astfel, „Vaip3s" semnifică „Verb auxiliar, indicativ, prezent, persoana a 3-a singular". Urmează numărul de identificare al regentului și funcția sintactică.

Formalismul în care lucrăm este numit gramatică de dependență, postulatele ei arată că o frază are un singur cuvânt independent, rădăcina, toate celelalte sunt subordonate de un alt element, fie că sunt cuvinte sau semne de punctuație. Acestea sunt nodurile arborelui, în vreme ce relația sintactică este inscripționată ca etichetă pe linia care marchează subordonarea. Acest tip de gramatică furnizează la fel de multă informație sintactică precum gramaticile de constituenți, dar permite o economie apreciabilă de noduri.

```
<sentence id="10_chat2_1_11" parser="Radu&apos;s parser" user="Ugla" date="2015-37-02">
 <word id="1" form="Asta" lemma="ăsta" postag="Pd3fsr" head="3" chunk="" deprel="sbj."/>
 <word id="2" form="a" lemma="avea" postag="Vaip3s" head="3" chunk="" deprel="aux."/>
 <word id="3" form="insemnat" lemma="insemnat" postag="Vmp" head="0" chunk=""/>
 <word id="4" form="k" lemma="că" postag="Cs" head="3" chunk="" deprel="pred."/>
 <word id="5" form="nu" lemma="nu" postag="Qz" head="8" chunk="" deprel="neg."/>
 <word id="6" form="ti" lemma="tu" postag="Pp2-sd--------w" head="8" chunk="" deprel="c.i."/>
 <word id="7" form="-a" lemma="avea" postag="Vaip3s" head="8" chunk="" deprel="aux."/>
 <word id="8" form="fost" lemma="fi" postag="Vmp" head="4" chunk="" deprel="subord."/>
 <word id="9" form="dor" lemma="dor" postag="Ncmsrn" head="8" chunk="" deprel="sbj."/>
 <word id="10" form="de" lemma="de" postag="Sp" head="9" chunk="" deprel="a.pron."/>
 <word id="11" form="noi" lemma="eu" postag="Pp1-pa--------s" head="10" chunk=""
deprel="prep."/>
 <word id="12" form="?" lemma="?" postag="QUEST" head="3" chunk="" deprel="punct."/>
 </sentence>
```

Figura 1. Arbore în format xml.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

Treebankul, numit UAIC-RoDepTb (*Treebank în dependency grammar* în limba română, constituit la Universitatea Al. I. Cuza) are peste 205.000 de cuvinte și semne de punctuație. Media este de aproximativ 19 cuvinte într-o frază sau propoziție. Este vorba de un treebank balansat, adică textele sunt (sau vor fi) selectate din toate stilurile funcționale ale limbii române. În acest an, au fost introduse două sub-corpusuri conținând texte de pe chat, totalizând 2.500 de fraze și propoziții. Alte sub-corpusuri conțin citate din DLR, texte regionale (PopReg), texte vechi (RoDia), texte legislative din *Acquis communitaire*, texte jurnalistice traduse din corpusul englez FrameNet, texte traduse din romanul *1984* de Orwell, texte beletristice românești. Treebankul conține minimum 1.000 de exemple din fiecare nou stil introdus, deoarece cu mai puține nu se pot antrena instrumentele de procesare automată. Fiind open source, poate fi utilizat în cercetările celor interesați[1].

1.2. *Particularități ale textelor Social Media*

Textele *Social Media* reprezintă o variantă non-standard de limbaj. Instrumentele de procesare cu care lucrăm sunt antrenate pe colecții foarte mari de limbaj standardizat, de aceea la procesarea textelor *social media* ele au înregistrat o scădere semnificativă a preciziei și a fost nevoie de antrenarea lor folosind un număr mare de astfel de texte nestandardizate, care nu respectă normele, pentru a obține o revenire a preciziei la valori comparabile cu ale unor programe similare. Acesta este motivul pentru care am fost nevoiți să procesăm și să corectăm manual nu câteva sute, ci câteva mii de exemple. Avontuur *et al.* (2012), Darling *et al.* (2012) și Toutanova *et al.* (2003) descriu de asemenea soluții pentru antrenarea instrumentelor de procesare a limbajului pe texte non-standard. În Liu *et al.* (2012) sunt descrise procedee automate pentru a „normaliza" textul *social-media*, pentru a-l aduce la o formă standard, ceea ce, în opinia noastră, reprezintă o ratare a șansei de a-l analiza în funcționalitatea lui, așa cum este.

Pentru a vizualiza forma arborescentă este nevoie de programe speciale. În figura 2 prezentăm un exemplu de frază simplă din treebank, vizualizată cu programul TreeAnnotator.

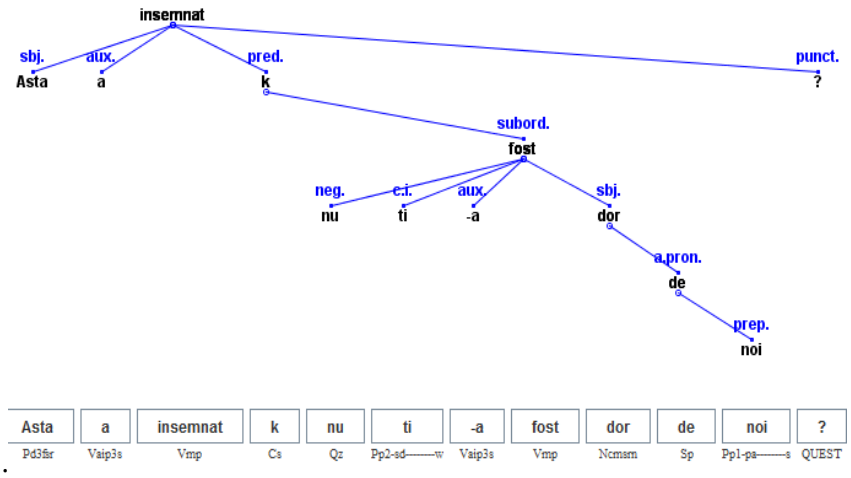


Figura 2. Fraza din Fig. 1 vizualizată în formă arborescentă cu programul Treeannotator.

[1] http://nlptools.info.uaic.ro/Resources.jsp/treebank_corpus.rar

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

POS-taggerul (adică analizorul morfologic automat) a fost antrenat atât pe texte cu litere cu semne diacritice românești (î, ș, ț, ă, â), cât și fără, fiindcă în comunicarea *Social Media* se găsesc ambele grafii; unii comunicatori folosesc litere cu diacritice, alții nu folosesc.

Am reuşit să supervizăm procesarea automată a celor 2.500 de fraze şi propoziţii şi deţinem acum un corpus pe care studiindu-l am făcut observaţii interesante despre tipul de încălcări ale standardelor lingvistice (prezente în alte stiluri ale limbii) pe care le găsim în comunicarea de pe chat.

Corpusul provine din copierea mesajelor transmise de persoane cu un înalt nivel cultural (deoarece la acelea am avut acces), astfel încât comunicările stereotipe și scurte alternează cu construcții lungi și elaborate, de obicei de tip narativ, dar uneori și de tip argumentativ.

S-ar putea studia măsura în care straturile cu nivel cultural diferit ale societății comunică între ele și se inter-influențează lingvistic prin intermediul *Social Media*. S-ar părea că persoanele cultivate împrumută tehnici (destul de inventive) de abreviere a comunicării (de pildă, în exemplul din Fig. 1-2 apare grafia „k” în loc de „că”), în vreme ce persoanele lipsite de imaginație împrumută figuri de limbaj de la persoanele cultivate și inventive, talentate (în corpus există replici ale unui scriitor și redactor șef de jurnal literar).

Acest tip de comunicare este în expansiune atât ca număr de utilizatori, cât și ca volum, tinde să devină un nou stil al limbii. Are particularităţi de tipul limbajului oral, dar şi jocuri de cuvinte sau abrevieri care utilizează forma scrisă a mesajelor; numeroase procedee sunt destinate prescurtării, rapidităţii transmiterii mesajelor.

Utilizează interferenţa stilurilor şi alte mijloace de factură ludică şi ironică. Nu respectă niciun fel de norme, cu excepția celor pragmatice, care asigură o bună comunicare între partenerii la schimbul de replici.

La nivelul discursului, am studiat şi schimburi între mai mult de doi parteneri, modul cum se intersectează mai multe teme şi cum se propun sau se refuză teme conversaţionale.

2. Particularități stilistice

Existența cuvintelor cu și fără semne diacritice dă naștere la numeroase ambiguități pe care comunicatorii le observă și le utilizează în mod ironic sau ludic. „Pana” poate și citit *până* sau *pană*, iar „putin” este interpretat nu drept „puțin”, adverb, ci drept numele propriu Putin. Abrevierile non-standard se bazează pe modul în care se pronunță literele izolate, neimplicate în cuvinte, c=ce, b=be, k=ca, iar în engleză p=pi, s=si, de unde:

(1) *C fk? (Ce faci?)*
(2) *psk (pisica).*

Cum începutul cuvântului este mai bogat în informație lexicală, iar finalul în informație gramaticală, referitoare la flexiune, se renunță la final. Alteori dispar silabe în interior ori se omit vocale, deoarece consoanele conțin mai multă informație decât vocalele:

(3) *cre k (cred că), dk (dacă)*
(4) *nik (nimica), adk (adică), aks (acasă)*

Remarcăm faptul că întotdeauna aceste comunicări își îndeplinesc funcția comunicativă, sunt decodate corect, cu toate că uneori destinatarul se preface că nu înțelege sau îl ironizează pe emitent pentru scrierea incorectă.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

În general, intenția ludică și cea de persiflare a conținutului serios al comunicării nu lipsesc decât din mesajele foarte scurte, cu intenția de a face un anunț sau care conțin o cerere urgentă de ajutor:

(5) *Titus nu mai vine, a intervenit ceva*.
(6) *Împrumută, te rooog, unul, de la prieteni, de unde poți!!!*

Chat-ul poate conține comunicare reală, cu intenții practice ori cu intenții confesive sau persuasive, însă de cele mai multe ori conținutul este (și) ludic, pretins serios – adică se admit două lecturi, existând un nucleu serios de comunicare și o serie de jocuri la nivelul expresiei. Uneori un șir de semne poate avea aerul unei greșeli de dactilografiere, dar conține aluzii ireverențioase (7, 10), pornografice (8, 11), contaminări între expresii antonime (9):

(7) *doar lucrez în cerșetare*
(8) *se mai întâmpulă*
(9) *mă aurzi?* (contaminare cu surzi, care nu aud)
(10) *electoralnic* (contaminare cu jalnic)
(11) *salutu-te*.

Utilizarea limbilor străine în moduri (voit) incorecte poate avea intenții ironice la adresa celorlalți vorbitori care o fac cu intenția de a epata:

(12) *forghes mi nacht* (Germ. vergiss mich nicht)
(13) *bonsoareta* (Fr. bonsoir)
(14) *Te prego* (It. prego).

În exemplul 12, expresia ar trebui să însemne „nu mă uita”, dar prin înlocuirea lui „nicht” cu „nacht” se obține sensul: „uită-mă noapte”.

O altă modalitate expresivă este nu doar parodierea stilurilor unor personaje, epoci, situații de comunicare, ci interferența acestor stiluri, așa încât rezultă expresii cu mari contraste, neadecvate cu niciunul dintre stiluri și cu niciuna dintre situații. În (15-16) se parodiază replici sau stilul unor scriitori (Șalom Aleichem, Caragiale, ambii umoriști):

(15) *Și șe fașeți, oi, oi, oi, cu atâtea? Miliție, pompieri, vidanjă?*
(16) *Bobocule, cum le știi tălică pe toate!*

Comentariul la o expresie poate fi ironic malițios, denaturând sensul, mimând neînțelegerea intenției emitentului:

(17) a: *eu nu am avut bursă de ieșit afară.*
b: *aha, bursă purgativă?*

Ironia la adresa celor care se lasă numiți de corespondenții lor „maestre” este exprimată prin diverse jocuri de cuvinte, aceștia sunt numiți „maeștri de ceremonii”, adică „ciocli, morticieni”, care „trag la pompe funegre”, iar femininul de la „mare maestru” este „mare măiastră”. Acest tip de comunicare este clar guvernată de intenția de a se amuza, poate defulând nemulțumiri cotidiene.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

3. Mărci pragmatice ale schimbului de replici

Studiind modalitățile de stabilire a contactului, de începere a schimbului de replici, observăm că acesta se poate face fie neutru, stereotip (1), fie în moduri ludice, inventive (18-19):

(18) *Săru-mânurile pă obraz.*
(19) a: *Ai înviat?*
b: *Adevărat am înviat!*

Verificarea existenței canalului de comunicare se face de asemenea în mod ludic (9) ori în mod convențional (20-21). Alte replici au funcția de a reclama rolul de emitent (22-23):

(20) *Ești on line?*
(21) *Ești acolo?*
(22) *Să-ți spun una* (*bună* ori *nebună*)!
(23) *Să-ți spun ceva în privat.*

Un schimb elaborat de replici are funcția de a propune o nouă temă de conversație, care este refuzată de interlocutor prin continuarea altei teme, iar apoi prin selecția ironic neadecvată a unei semnificații mitice pentru un personaj banal:

(24) a: *Tu îți amintești de unul, Florin Toma?*
(4 replici pe altă temă)
a: *Dar de Toma, ce zici?*
b: *care, necredinciosul?*
a: *Nu, Florin!*

4. Analiza discursului

4.1. *Teme discursive interțesute cu doi interlocutori*

Întrepătrunderea și încurcarea între ele a firelor discursive este în parte cauzată de viteza redusă de scriere a unor participanți la schimbul de replici. În timp ce scrie, unul dintre participanți nu poate urmări ceea ce scrie celălalt și acesta începe o altă temă la care interlocutorul este silit să răspundă cu întârziere. Fenomenul nu se produce în conversațiile orale. Pentru cazurile ambigue, când o replică ar putea fi legată de ambele fire tematice, putem folosi spre dezambiguizare fie analiza atitudinilor, fie teoria nervurilor (Cristea 2005a). Exemplu:

(25) (B) 805: Dar de ce? Nu vrea frumușelul să îl văd?
(C) 806: Azi am scos pisica afară de două ori.
(B) 807: E la un coleg, socializează.
(C) 808: Era foarte speriată, s-a dus sub smochin.
(B) 809: A venit acasă, și-a lăsat ghiozdanul și a plecat.
(?) 810: Aseară tot așa a făcut.
(?) 811: Aha, cred că se va obișnui în timp.

Se găsesc aici două teme: refuzul tânărului de a face o vizită (B) și educația igienică a pisicii (C). Problema este că ultimele două replici ar putea fi atașate în mod eronat sau ironic la tema (B), deoarece nu au mărci ale legăturii cu tema (C).

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

În NLP se încearcă găsirea unor algoritmi care să explice în ce mod inteligența umană sesizează cu ușurință faptul că e vorba despre o anumită temă în situații ambigue, și apoi introducerea acestor algoritmi în programele de procesare a limbajului natural.

Dacă legăm greșit replicile de mai sus, putem obține interpretări absurd-amuzante:

(26) pisica e la o colegă, socializează.
(27) pisica și-a lăsat ghiozdanul și s-a dus sub smochin.
(28) tânărul se va obișnui în timp să plece de acasă în fiecare noapte.

Putem soluționa ambiguitatea în termenii analizei sentimentelor (atitudinilor). Termenul „socializează” (807) este selectat ironic. El aparține sferei relațiilor oficiale și nu celor familiale. Selecția inadecvat-ironică este o marcă a dezaprobării, evaluării negative a acțiunii tânărului de către povestitoare, pe când (811) este evaluată pozitiv, obișnuirea este sperată, deci dezirabilă. Evaluarea fiind diferită, cele două replici nu fac parte din același fir discursiv (Fig. 3).

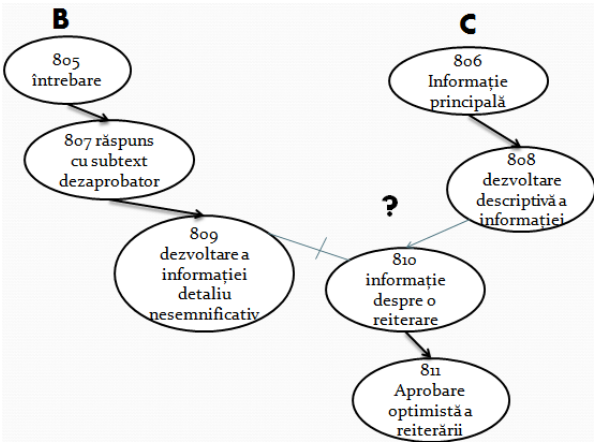


Figura 3. Arbore discursiv al replicilor din exemplul 25.

Desigur, nu este vorba de un discurs, ci de două fragmente de discurs, ceea ce se remarcă în primul rând datorită nesaturării informative la limitele fiecărei serii de replici. În arborele B firul principal este o întrebare despre băiat (de ce nu vrea să ne vedem azi), care reclamă un răspuns, fără a primi unul foarte clar. Sunt furnizate detalii irelevante care trebuie interpretate (negativ, pentru că acțiunile băiatului sunt prezentate ca aberante: să „socializeze” în timpul liber cu colegii, la ore foarte târzii).

Arborele C conține o informație nouă și o dezvoltare cu detalii descriptive.

Arborele final conține o informație despre reiterarea unei acțiuni și o evaluare pozitivă. Nu există mărci ale genului gramatical al eroului reiterării: dacă ar fi vorba de masculin, ar fi personajul din arborele B; dacă ar fi vorba de feminin, de personajul din arborele C (tot așa a făcut el sau ea).

Reiterarea nu constituie un răspuns la întrebarea care formează firul principal al arborelui B. De ce nu vrea să mă vadă, să îl văd (astă seară)? ≠ Aseară tot așa a făcut – este incoerent (nimeni nu a întrebat de ce nu a vrut aseară să fie văzut, întrebarea este la timpul prezent). Legarea de informațiile redundante din 807 și 809 ar mări cantitatea de detalii aberante și inutile.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

În schimb, reiterarea este o continuare firească a informației din arborele C: informație – descriere, nouă acțiune, consecință a descrierii (speriată și de aceea a fugit sub smochin) – generalizare prin prezentarea acțiunii ca o reiterare a uneia trecute. Evaluarea finală este pozitivă, deci nu concordă cu calificarea drept aberant a comportamentului din arborele B: De ce ... ? Răspuns implicit: Nu există motiv valabil pentru care să nu îl poți vedea pe băiatul frumușel (dar, de fapt, el nu e disponibil pentru a fi văzut, fiind beat, informație disimulată).

Arborele C descrie comportamentul pisicii pe cale să se obișnuiască să iasă afară pentru a-și face necesitățile, deci reiterarea acțiunii este evaluată pozitiv. Reiterarea unui șir de acțiuni prezentate ca absurde ori negative în arborele B ar fi incoerent să fie evaluată pozitiv; prin acumulare, caracterul negativ ar trebui să crească.

4.2. *Teme discursive în conversația mai multor persoane*

În figura 4 sunt trasate relații între replicile a patru interlocutori care comentează o fotografie veche postată de unul dintre ei. Firele se îndepărtează tot mai mult de pretextul începerii conversației între persoane aflate în spații geografice foarte îndepărtate. Acest fapt are drept cauză faptul că fotografia este un pretext de a declanșa comunicarea la distanță, plăcerea comunicării la distanță între persoane cândva apropiate fiind adevăratul mobil al schimburilor de replici:

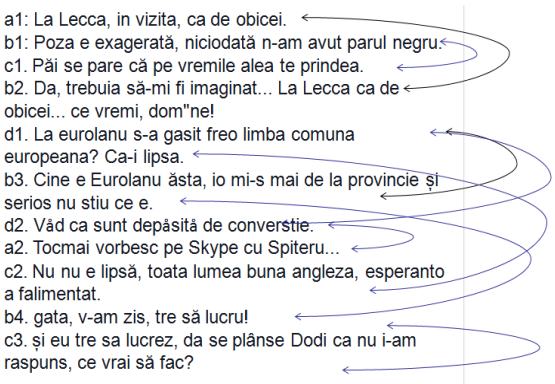


Figura 4. Schimb de replici între 4 interlocutori.

Fragmentul citat este deschis în partea superioară, insuficient informațional, conținând trimiteri la replici anterioare, și închis în partea finală, conținând mărci de închidere și saturare informațională (relativă, fără răspuns la întrebarea b3).

Observăm că există patru teme diferite. Prima dintre ele, fotografia, nu are nicio relație cu celelalte. Se propune apoi tema Eurolan, care trimite la o postare anterioară, pentru că este vorba de fragmentul final dintr-un șir mai lung de replici și postări. Sunt puse două întrebări dintre care prima nu primește niciun răspuns, iar a doua primește unul doar în urma unui protest, care reprezintă a treia temă. Impresia este de dialog al surzilor, în care nimeni nu vrea să răspundă la întrebările celorlalți, ceea ce pare să-i distreze pe participanți, cu excepția celui care protestează. O altă temă ar fi munca sau, altfel spus, invocarea unor motive pentru încetarea conversației, ceea ce ar include și replica a2, care invocă o altă conversație.

Impresia de dialog al surzilor este dată și de existența a 4 teme diferite în 11 replici, cu izolarea completă a primeia dintre ele, după ce acumulează patru intervenții, astfel încât următoarele 3 teme se consumă în 7 intervenții cu întortocheate și superficiale relații între ele.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

Interlocutorul a, care postează fotografia și explică (răspunzând la o replică anterioară) care este locul unde a fost făcută, se retrage imediat sub pretextul altei conversații (deși practica desfășurării în paralel a mai multor conversații este una obișnuită).

Interlocutorii b și c invocă necesitatea întoarcerii la lucru pentru a justifica întreruperea conversației, ceea ce reprezintă tema 4 (vai, cât suntem de ocupați).

Schema tematică a replicilor din Fig. 4 este redată în fig. 5.

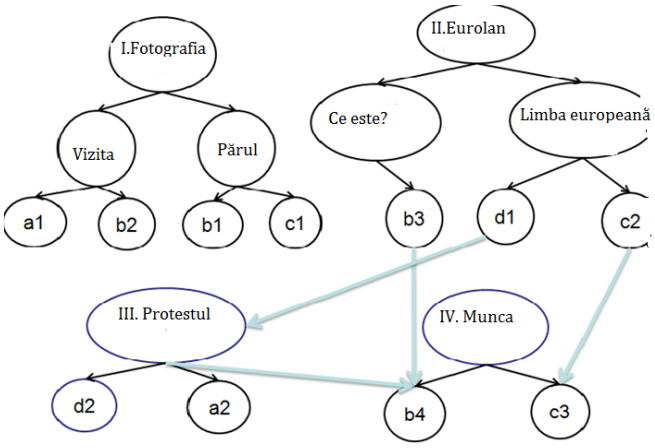


Figura 5. Structura tematică a replicilor din fig. 4.

S-ar mai putea găsi și alte teme izolate, cum ar fi nostalgia („ce vremi, dom'ne"), declanșată de poza veche; tema evocării unui moment temporal trecut apare și în replica c1 („pe vremile alea"), dar replica primește o evaluare negativă în b1 (vremi = niciodată).

În Cristea (2005b) găsim o teorie a adăugării de informație nouă la frontiera dreaptă a arborelui discursiv. În figura 5 se poate observa că informația din tema III e adăugată în arbore la frontiera stângă, pentru că este declanșată de d1, care este o replică anterioară lui c2, declanșatoare a temei IV, aceasta din urmă fiind adăugată într-adevăr pe frontiera dreaptă a arborelui. Arborele discursiv interțesut încalcă adeseori teoria frontierei din dreapta. Nici la nivelul arborelui unei propoziții teoria nu este respectată de limbile cu o topică liberă, cum este româna.

În ce privește stilul conversației de mai sus, observăm mimetisme care mărturisesc clar intenția ludică: „să lucru" (regionalism Transilvania), „se plânse" (regionalism Oltenia), „vrai", „vremi" (arhaisme), „dom'ne" (caragialism, adresat unor femei), „io mi-s mai de la provincie" (limbaj popular, vorbit, emitenta e din Santa Fé).

5. Concluzii

Mediul de comunicare prin internet este lipsit atât de limite spațiale, cât și de limite convenționale, propune o lume virtuală în care totul este permis din punct de vedere stilistic și pragmatic. Poate revendica locul de principal laborator al inventivității, inovației lingvistice, pe care îl deținea până în prezent stilul oral.

Studiul acestui tip de comunicare este atât necesar, cât și interesant; fără acest studiu în curând nu o să ne putem explica direcția în care evoluează limba, atât formal, cât și ca mijloace de expresie, de comunicare.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)

Prin procesarea automată a exemplelor se poate obține o cantitate mult mai mare de date lingvistice care pot fi supuse interpretării sau pot oferi argumente statistice pentru susținerea unei teorii.

Diferite teorii și formalisme gramaticale sunt ușor de tradus unul în termenii altuia; gramatica transformațională, gramatica de constituenți nu este unica posibilitate de a analiza lingvistic datele colectate din comunicarea în limbajul natural.

Este necesară o mai mare deschidere a cercetătorilor în domeniul lingvisticii generale și a celor din domeniul lingvisticii computaționale unii către contribuțiile celorlalți.

BIBLIOGRAFIE

Avontuur, Tetske, Iris Balemans, Laura Elshof, Nanne van Noord, Menno van Zaanen, 2012, „Developing a part-of-speech tagger for Dutch tweets", *Computational Linguistics in the Netherlands*, 2, p. 34–51.

Cristea, Dan, 2005a, „Motivations and Implications of Veins Theory", în Bernadette Sharp (ed.), *Natural Language Understanding and Cognitive Science*, *Proceedings of the 2nd International Workshop on Natural Language Understanding and Cognitive Science*, Miami, U.S.A., May 2005, INSTICC Press, Portugal, p. 32–44.

Cristea, Dan, 2005b, „The Right Frontier Constraint Holds Unconditionally", *Proceedings of the Multidisciplinary Approaches to Discourse*, Chorin/Berlin, Germany.

Darling, M. William, Michael J. Paul, Fei Song, 2012, „Unsupervised part-of-speech tagging in noisy and esoteric domains with a syntactic-semantic Bayesian HMM", *Proceedings of the conference of the European chapter of the Association for Computational Linguistics*, p. 1–9.

Liu, Fei, Fuliang Weng, Xiao Jiang, 2012, „A broad-coverage normalization system for social media language", *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, p. 1035–1044.

Mărănduc, Cătălina, Cenel-Augusto Perez, 2015, „A Romanian Dependency Treebank", *International Journal of Computational Linguistics and Applications*, vol. 6, no. 2, issue July-December 2015, p. 25–40.

Perez, Cenel Augusto, 2015, „Un corpus de texte pentru limba română adnotat sintactic sub formă de arbori", în Rodica Zafiu, Isabela Nedelcu (eds.), *Actele celui de-al 14-lea colocviu internațional al Departamentului de lingvistică al Facultății de Litere, Universitatea București*, vol. I, București, Editura Universității din București, p. 103–113.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, Yoram Singer, 2003, „Feature-rich part-of-speech tagging with a cyclic dependency network", *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, p. 173–180.

STYLISTIC AND PRAGMATIC PARTICULARITIES OF SOCIAL MEDIA TEXTS

(*Abstract*)

In this article we have analysed the spelling, lexical, pragmatic and discursive peculiarities of social media texts that use a non-standardized language, free and creative, making their processing very difficult with computerized instruments. It is a language whose weight and importance is growing, and certain phenomena of contemporary Romanian evolution cannot be understood without the study of social media texts.

BDD-V2330 © 2016 Editura Universității din București
Provided by Diacronia.ro for IP 216.73.217.75 (2026-11-06 20:57:10 UTC)