

# UTILIZAREA TEHNOLOGIILOR DE PRELUCRARE A LIMBAJULUI NATURAL ÎN CONTEXT ELECTORAL: METODA LIWC-2007<sup>1</sup>

Cercet. post-doc Daniela GÎFU  
Universitatea „Alexandru Ioan Cuza”, Iași

## Abstract

This paper presents and interprets the 2009 presidential campaign through a new computational technique of political discourse linguistic processing (method LIWC-2007 - Linguistic Inquiry and Word Count), providing a rapid and robust possibility to understand the public sphere through political language. The concept behind this method is that the manner in which individuals speak and write represents an open window to their emotional and cognitive worlds. In other words, the vocabulary identifies the emotionalism of a speech. Investigation supports research specialists in political science, being helpful, mainly in the attached social prospects of election campaigns, to measure media reactions of the developments in the political arena.

**Keywords:** political speech, computational technique, linguistic analysis, political language, electoral context

## 1. Introducere

Opțiunea pentru un asemenea studiu - prezentarea și interpretarea contextului electoral cu ajutorul tehnicii computaționale de prelucrare lingvistică a discursurilor politice (metoda LIWC-2007 - *Linguistic Inquiry and Word Count* -, dezvoltată în colaborare de Universitatea din Texas at Austin și Universitatea din Auckland), a evidențiat influența discursului politic asupra receptorilor săi, discursivitatea politică instituind o realitate posibilă și alternativă.

Metoda LIWC-2007, a fost aplicată pentru prima dată pentru limba română pe un corpus discursiv-electoral la finele anului 2009, rezultatele fiind prezentate în luna martie a.c.<sup>2</sup>, oferă o posibilitate rapidă și riguroasă de a înțelege sfera publică prin intermediul limbajului, mai precis al discursului politic, situat la intersecția a trei spații simbolice importante: spațiul politic, spațiul public și spațiul comunicațional<sup>3</sup>. Limbajul politic, esențial în construirea discursului<sup>4</sup>, necesită o abordare interdisciplinară<sup>5</sup> în care lingvistica conlucrează cu retorica, cu științele comunicării, cu politologia și cu sociologia. Folosirea

---

<sup>1</sup> Această lucrare răspunde tematicii proiectului de cercetare științifică postdoctorală, propus în iunie 2010, din cadrul Departamentului de cercetare științifică al *Facultății de Informatică și al Facultății de Psihologie și Științe ale Educației*, domeniul *D4 (Științe ale omului și ale societății)*, proiect finanțat prin Programul Operațional Sectorial Dezvoltarea Resurselor Umane, proiectului POSDRU/89/1.5/S/63663: „*Rețea transnațională de management integrat al cercetării postdoctorale în domeniul Comunicarea științei. Construcție instituțională (școală doctorală) și program de burse (CommScie)*”.

<sup>2</sup> Daniela GÎFU, *Discursul presei scrise și violența simbolică. Analiza unei campanii electorale*, teză de doctorat susținută la Facultatea de Filosofie, Universitatea „Al. I. Cuza”, Iași, 19 martie 2010  
(v. [http://www.uaic.ro/uaic/bin/download/Academic/Doctorate\\_martie\\_2010/GfuC.Daniela.pdf](http://www.uaic.ro/uaic/bin/download/Academic/Doctorate_martie_2010/GfuC.Daniela.pdf))

<sup>3</sup> Dominique WOLTON, *Les contradictions de la communication politique*, în Philippe CABIN (ed.) *La communication. État des savoirs*, Éd. Sciences Humaines, Paris, 1998, pp. 344-345

<sup>4</sup> David D. PERLMUTTER, *The Manship School guide to political communication*, Baton Rouge, Louisiana State University Press, 1999, în Călin SINESCU, *Comunicare politică*, Ed. Universitară, București, 2007, pp. 166-167

<sup>5</sup> Laurențiu ȘOITU, *Philosophy of communication*, Ed. Institutul European, Iasi, 2001

limbajului în politică are rolul de a „sanctifica acțiunea”<sup>6</sup>, câștigând acordul deplin, de durată, al electoratului.

## 2. Metodologia de lucru

În elaborarea concluziilor preliminare privind desfășurarea procesului electoral prezidențial, realizat în perioada octombrie-decembrie 2009, am cules, stocat și centralizat electronic în două etape (turul I și turul II al campaniei prezidențiale) texte politice, editorialele a 3 publicații naționale cu profil comun<sup>7</sup>, *Evenimentul zilei*, *Gândul* și *Ziua*, precum și discursurile politice<sup>8</sup> (în ambele forme: oratorică și dialogică) a celor patru prezidențiabili cotați<sup>9</sup> pe primele patru locuri: Crin Antonescu (PNL), Traian Băsescu (PDL), Mircea Geoană (PSD+PC) și Sorin Oprescu (independent).

### 2.1. Softul LIWC-2007

În componența programului intră un dicționar, care, în varianta originală americană, include 4500 de cuvinte și rădăcini de cuvinte. În varianta românească acesta are aproximativ 5100 de cuvinte și rădăcini (diferența în plus datorându-se cu precădere mai marii productivități lexicale a limbii române prin flexionare decât a limbii engleze, care a trebuit oglindită prin includerea mai multor rădăcini pentru același cuvânt). Rădăcinile sunt marcate prin simboluri terminate prin semnul \*.

Fiecărei intrări de dicționar (cuvânt sau rădăcină) îi este asignată o listă de categorii semantice. Programul incrementează contoarele corespunzătoare claselor respective la fiecare detectare a aceluși cuvânt (ori a acelei rădăcini) în textul de intrare. De exemplu, rădăcinii *plictico*\* îi sunt asignate clasele: 125=emoțional, 127=emoțional\_negativ și 130=supărare. De fiecare dată când acest cuvânt ori o variantă a lui este detectată în fișierul de intrare, toate cele 3 contoare, corespunzătoare claselor menționate, sunt incrementate.

Există o parțială clasificare ierarhică a claselor. De exemplu, orice cuvânt care aparține clasei **supărare** aparține și clasei **emoțional\_negativ** și toate cuvintele acestei clase aparțin și clasei **emoțional**. În varianta originală, LIWC numără 64 de categorii (sintactice și semantice)<sup>10</sup>, din care noi am reținut 28 de clase, considerate cu un grad optim în

<sup>6</sup> Murray EDELMAN, *Politica și utilizarea simbolurilor*, Ed. Polirom, Iași, 1999, p. 111

<sup>7</sup> Aceste trei publicații au *profil comun*, cf. MediaPress ([www.mediapres.ro](http://www.mediapres.ro)): sunt cotidiene naționale de informație generală, se prezintă sub formă de tabloide, cu un tiraj de la câteva zeci de mii de exemplare pe ediție, forma tipărită a editorialelor se regăsește integral și pe paginile de internet ale celor trei cotidiene monitorizate: *Evenimentul zilei* – [www.evz.ro](http://www.evz.ro), *Gândul* – [www.gandul.info](http://www.gandul.info), *Ziua* – [www.ziua.ro](http://www.ziua.ro).

<sup>8</sup> Surse de informare: *Crin Antonescu* – [www.crinantonescu.ro](http://www.crinantonescu.ro), [www.pnl.ro](http://www.pnl.ro), *Traian Băsescu* – [www.basescu.ro](http://www.basescu.ro), [www.pdl.org.ro](http://www.pdl.org.ro), [www.presidency.ro](http://www.presidency.ro), *Mircea Geoană* – [www.mirceageoana.ro](http://www.mirceageoana.ro), [www.psd.ro](http://www.psd.ro) și *Sorin Oprescu* – [www.sorinoprescu.ro](http://www.sorinoprescu.ro).

<sup>9</sup> Amintim sondajul privind alegerile prezidențiale, publicat în 12 octombrie 2009. Acesta a fost realizat în perioada 17 septembrie - 7 octombrie de CSOP, pe un eșantion național de 18.348 de persoane, reprezentativ pentru populația în vârstă de 18 ani și peste. Eroarea este de ±0,7%, iar probabilitatea de 0,95. Intenția de vot: Traian Băsescu - 37%, Mircea Geoană - 24%, Crin Antonescu - 13%, Sorin Oprescu - 10%.

<sup>10</sup> Amintim cele 64 de clase, clasificate în 4 categorii, precedate de un anumit cod regăsit în structura programului LIWC2007: *lingvistică* (22 de clase) - 1 funct (Total function words), 2 pronoun (Total pronouns), 3 ppron (Personal pronouns), 4 i (1st pers singular), 5 we (1st pers plural), 6 you (2nd person), 7 shehe (3rd pers singular), 8 they (3rd pers plural), 9 ipron (Impersonal pronouns), 10 article (Articles), 11 verb (Common verbs), 12 auxverb (Auxiliary verbs), 13 past (Past tense), 14 present (Present tense), 15 future (Future tense), 16 adverb (Adverbs), 17 preps (Prepositions), 18 conj (Conjunctions), 19 negate (Negations), 20 quant (Quantifiers), 21 number (Numbers), 22 swear (Swear words); *psihologică* (32 clase) - 121 social (Social processes), 122 family

determinarea unei anumite *atitudini politice* care poate influența decizia de vot. Raportându-ne la criteriul *convingere*, pe baza căruia se construiește fiecare etapă a discursului politic, finalitatea urmărită va evidenția legătura între proporția claselor predominante în discurs și opțiunea de vot. Menționăm că au fost păstrate codurile claselor ca în programul american, numele clasei fiind tradus în limba română.

Tabel 1. Descrierea celor 28 de clase reținute în varianta românească a softului LIWC-2007

Cod	Clasa eng. (abr.)	Clasa rom.	Descrierea semnificației clasei în funcție de lexicurile atașate corespunzător
22	swear	injurii	Cuvinte care lezează demnitatea/reputația cuiva. Exemple: <i>afurisit, cretin, jigodie, nemernic, nevrednic...</i>
121	social	social	Cuvinte cu referire la apartenența la un anumit grup social. Exemple: <i>celebrare, președinție, consultație, eveniment, intervenție...</i>
122	family	familie	Exemple: <i>clan, fii, înrudire, națiune, părinți.</i>
123	friend	prieteni	Exemple: <i>camarazi, colegi, companioni, membru, partener...</i>
124	human	oameni	Exemple: <i>adult, individ, orășeni, personalitate, tipi...</i>
125	affect	emoțional	Cuvinte care provoacă emoții (pozitive sau/și negative), transpun realitatea în zona afectivă. Exemple: <i>abătut, binevoitor, curajos, impasibil, necinstit, revigorant...</i>
126	posemo	emoțional_pozitiv	Cuvinte care provoacă emoții pozitive. Exemple: <i>ademenitor, fermecător, plăcut, tonic, popular...</i>
127	negemo	emoțional_negativ	Cuvinte care provoacă emoții negative. Regăsim o parțială clasificare ierarhică - clasele: neliniște, furie, supărare. Exemple: <i>agresiv, cinic, obscen, pungaș, trândav...</i>
128	anx	neliniște	Exemple: <i>alarmant, bănuitor, nervos, descumpănitor, tensionat, vulnerabil, refractar.</i>
129	anger	furie	Exemple: <i>agitat, brutal, irascibil, mincinos, ofensator...</i>
130	sad	supărare	Exemple: <i>îngrijorător, plicticos, revoltător, trist, pesimist...</i>
131	cogmech	rațional	Cuvinte care exprimă un raționament, o judecată. Exemple: <i>bui măcit, discernământ, inventiv, laș, precaut...</i>
132	insight	intuiție	Exemple: <i>ager, explicativ, lămuritor, logic, realist...</i>
133	cause	determinare	Exemple: <i>afectat, deductiv, independent, insinuant, manipulator...</i>
135	tentat	nesiguranță	Exemple: <i>ambiguu, confuz, ipotetic, probabilistic, vag...</i>
136	certain	siguranță	Exemple: <i>categoric, clar, exact, hotărât, just...</i>
137	inhib	inhibiție	Exemple: <i>absent, blocat, descurajat, prevăzător, șovăitor...</i>
140	percept	perceptiv	Cuvinte în legătură cu facultatea de a percepe vizual, auditiv și tactil. Exemple: <i>aude, consonant, febril, palpabil, observabil...</i>
141	see	vizual	Cuvinte în legătură cu facultatea de a percepe vizual.

(Family), 123 friend (Friends), 124 humans (Humans), 125 affect (Affective processes), 126 posemo (Positive emotion), 127 negemo (Negative emotion), 128 anx (Anxiety), 129 anger (Anger), 130 sad (Sadness), 131 cogmech (Cognitive processes), 132 insight (Insight), 133 cause (Causation), 134 discrep (Discrepancy), 135 tentat (Tentative), 136 certain (Certainty), 137 inhib (Inhibition), 138 incl (Inclusive), 139 excl (Exclusive), 140 percept (Perceptual processes), 141 see (See), 142 hear (Hear), 143 feel (Feel), 146 bio (Biological processes), 147 body (Body), 148 health (Health), 149 sexual (Sexual), 150 ingest (Ingestion), 250 relativ (Relativity), 251 motion (Motion), 252 space (Space), 253 time (Time); *preocupări socio-profesionale* (7 clase) - 354 work (Work), 355 achieve (Achievement), 356 leisure (Leisure), 357 home (Home), 358 money (Money), 359 relig (Religion), 360 death (Death) și *paralingvistică* (3 clase): 462 assent (Assent), 463 nonfl (Nonfluencies), 464 filler (Fillers).

			Exemple: <i>holbat, licărire, lumină, orb, scânteie...</i>
142	hear	auditiv	Cuvinte în legătură cu facultatea de a percepe auditiv. Exemple: <i>acustic, consonant, răspicat, tăcut, surd...</i>
143	feel	tactil	Cuvinte în legătură cu facultatea de a percepe tactil. Exemple: <i>ascuțit, fragil, estompat, îndemânat, neputincios...</i>
149	sexual	sexual	Cuvinte cu conotație sexuală. Exemple: <i>atingere, dezmiardare, încins, mângâie, pipăie...</i>
354	work	muncă	Cuvinte cu referire la activități/instituții specifice muncii. Exemple: <i>calificare, organizație, departament, privilegii, proiecte...</i>
355	achieve	realizări	Cuvinte cu referire la acțiuni de eficientizare a vieții socio-umane. Exemple: <i>competențe, achiziții, facultăți, prelegere, succese...</i>
356	leisure	agrement	Cuvinte cu referire la activitățile de recreere. Exemple: <i>artist, camping, cazinou, dans, whisky...</i>
357	home	cămin	Cuvinte cu referire la arealul căminului. Exemple: <i>cameră, clădire, decor, mansardă, vatră...</i>
358	money	bani	Cuvinte referitoare la domeniul economico-financiar. Exemple: <i>afacere, avere, conturi, financiar, licitații...</i>
359	relig	religie	Cuvinte cu referire la zona spirituală. Exemple: <i>Biblie, Dumnezeu, cruce, creștini, milostenie...</i>

Au fost eliminate clasele care dau o categorizare asupra părții de vorbire a cuvintelor (pronume, verbe, adverbe, prepoziții, conjuncții etc.) și implicit asupra timpurilor verbelor (prezent, trecut, viitor), precum și câteva clase din categorii precum: procese cognitive (inclusiv, exclusiv), procese biologice (corp, sănătate, ingestie), relativitate (spațiu, timp, mișcare) etc. În esență softul realizează o numărare a apariției cuvintelor corespunzătoare diferitelor clase semantice, întorcând tabele de frecvențe. Acestea pot fi apoi ușor transformate în grafice.

## 2.2 Alegerea lexicului

Alegerea lexicului englezesc al LIWC-2007 (și a variantei precedente LIWC2001<sup>11</sup>) s-a făcut în mai mulți pași (în perioada anilor 1992-2007):

- s-a făcut mai întâi o selecție, folosind ca surse dicționare (Roget's Thesaurus ș.a.), urmată de sesiuni de brain-storming în care dimensiunile claselor au fost mărite;
- colecția astfel obținută a fost apoi trecută în mâinile a 3 experți, care au lucrat independent unul de altul. Ei au avut misiunea să decidă asupra fiecărui cuvânt, dacă trebuia ori nu păstrat în categoria respectivă, și au avut libertatea de a adăuga cuvinte noi. S-a aplicat apoi un sistem de votare, pentru a decide pe bază de majoritate care cuvinte să fie păstrate;
- s-a făcut apoi o evaluare pe un corpus de texte, în urma căreia un număr de clase, a căror reprezentativitate s-a dovedit slabă (mai puțin de 0.3%), au fost eliminate, după cum au fost eliminate și cuvinte a căror rată de apariție în clase a fost mai mică de 0.005%;

<sup>11</sup> Documentația originală a LIWC menționează testele de validitate care au însoțit darea în exploatare a programului. Se apreciază că aproximativ 86% din lexicul utilizat curent de indivizii umani în scris ori în vorbire se regăsește în LIWC. Testul s-a făcut pe un corpus englezesc (cules în SUA, Canada și Noua Zeelandă) de aproximativ 167 milioane de cuvinte, incluzând registrele: scrieri emoționale, scrieri neutre din punct de vedere emoțional, știință, beletristică, blogg-uri și transcrierea vorbirii, pe o perioadă care se întinde aproximativ între anii 1700 și 2007. Unul dintre primele teste de validare ale LIWC a fost efectuat de Pennebaker și Francis în 1996. (v. James W. PENNEBAKER, Martha E. FRANCIS, Roger J. BOOTH, *Linguistic Inquiry and Word Count – LIWC2001*, Erlbaum Publishers, Mahwah, NJ, 2001 și pe [www.erlbaum.com](http://www.erlbaum.com)).

- varianta actuală (2007) a dicționarului a suferit modificări substanțiale față de varianta 2001. Printre altele, o echipă de 4 experți a evaluat care cuvinte, din fondul celor mai frecvente 2000 de cuvinte ale limbii engleze, ar trebui incluse în dicționarul LIWC.

Dicționarul LIWC a fost tradus (după originarul englezesc) în câteva limbi, printre care: spaniolă, italiană, germană, olandeză, norvegiană, portugheză etc.

### 2.3 Construirea lexicului românesc

Ținând cont de procesul extrem de sofisticat și de durată după care s-a constituit dicționarul LIWC-2007, pentru utilizarea în cercetarea descrisă în această lucrare am decis asupra unui proces de localizare prin traducere, în defavoarea unuia în care colectarea dicționarului ar fi fost complet decuplată de varianta englezească. Desigur, am fost conștienți de riscul major pe care o astfel de decizie l-ar putea avea și care constă în obținerea unui lexic puternic influențat de limba engleză, prin preluarea termenilor (echivalenților de traducere) din dicționar, în afara contextelor semantice, cât și invers, prin neinclusiunea unor termeni curent utilizați în limba română, dar care nu sunt sugerați imediat de echivalenții originali. Pentru a corecta aceste dezavantaje, procesul de constituire a dicționarului românesc s-a desfășurat în următoarele etape:

- mai întâi termenii englezești s-au tradus, în ordine alfabetică (cu preluarea identică a claselor). În această etapă s-a decis, pentru fiecare cuvânt ori rădăcină inclusă, care trebuie să fie familia de rădăcini a lui, capabilă să asigure o acoperire cât mai bună a variabilității flexionare, fără însă a capta (prin procesul de *pattern-matching*) termeni străini de familia avută în vedere;
- lexicul astfel constituit a fost apoi sortat după fiecare număr de clasă, obținându-se un număr de 64 de liste (evident, nedisjuncte, datorită posibilității ca anumiți termeni să aparțină la mai multe clase);
- s-au reținut din acestea doar 28 de clase, considerate semnificative pentru acest tip de analiză.
- cuvintele/rădăcinile din fiecare clasă au fost apoi sortate alfabetic;
- s-a evaluat apoi fiecare clasă în parte, eliminându-se cuvintele care puteau introduce ambiguități și incluzând alte cuvinte/rădăcini din sfera sinonimică a celor existente. Acest proces a fost ușurat de faptul că s-a lucrat asupra unei clase (sfere semantice) odată. Corectarea în interiorul aceleiași clase are avantajul că ochiul uman poate urmări doar cuvinte din aceeași sferă semantică și poate astfel recunoaște ușor când au fost făcute greșeli de clasificare. Totodată, ordonarea alfabetică oferă posibilitatea operării anumitor optimizări de scriere a familiilor de cuvinte, prin exploatarea jolly-jokerului \* (semnul care marchează apariția a zero ori mai multe caractere la sfârșitul rădăcinii unui cuvânt)<sup>12</sup>;
- în final, toate dicționarele asignate claselor au fost puse împreună, lista astfel obținută s-a sortat din nou alfabetic și s-au eliminat aparițiile multiple (prin concatenarea listelor de clase, cu lăsarea o singură dată a unui număr de clasă în lista unui cuvânt/unei rădăcini).

---

<sup>12</sup> În informatică, simboluri de forma xxx\*, unde xxx reprezintă un număr oarecare de caractere, se numesc expresii regulate.

## 2.4 Optimizări în dicționar

În faza corectării dicționarelor asignate claselor s-au operat o seamă de optimizări asupra cuvintelor. Iată o listă de exemple:

### Condensări:

modifica\*, modificat\*, modifică\*, modificând  
înlocuite cu: *modific\**

care are și avantajul de a include și formele: *modific*, *modifici*  
ce ar fi fost ignorate în varianta precedentă.

### Grupuri ce erau cuprinse unul în altul, precum:

necesi\*, necesit\*  
au fost reduse la *necesi\**

sau grupul: *ascul\**, *ascult*, *ascult\**, *ascultare\**, *ascultat*, *ascultați*, *ascultă*, *ascultăm*, *ascultări\**,  
*ascultăto\**, *asculți*  
a fost redus la *ascul\**

### Completarea cu forme noi:

*noutăți\**  
a fost înlocuit cu: *noutate\**, *noutăți\**

### Cuvinte ambigue eliminate:

*însuș\** la clasa 133 (**determinare**), cu intenția de a genera familia verbului *a însuși*, pentru că poate genera coliziuni cu *însuși* a fost transformat în *însușit\**.

Intrarea *prefăcut\**, formă a verbului *preface*, a fost eliminată din clasa 133 (**determinare**), pentru că generează confuzia cu sensul reflexiv (care este un sens emoțional\_negativ).

### Eliminarea unor cuvinte aflate în clase neadecvate:

*grăsim\**, eliminat din clasa 140 (**perceptiv**), pentru că deși grăsimea poate fi un aliment și deci e legat de gust, deci de percepție, ea poate avea și sensul de stare a unei ființe grase.

### Alte observații

În clasa 355 (**realizări**) sunt incluse atât cuvinte care numesc realizare / succes / încununare (majoritatea) cât și unele care dau nume înfrângerii / nerealizării / nereușitei.

*Exemple:* *mediocr\**, *neadecva\**, *necalifica\**, *necorespunzăto\**, *neîndestulăto\**, *nepricepu\**, *neputincio\**, *nereușit\**, *nesăbui\**, *renunț\**, *rata\**.

## 3. Studiu de caz: Alegerile prezidențiale 2009

Întregul corpus discursiv selectat din presa scrisă monitorizată a fost procesat cu softul LIWC-2007 (al cărui parametru este calculul de frecvență - analiza cantitativă), rezultatele fiind apoi interpretate calitativ (analiza calitativă) pe baza reprezentărilor grafice realizate cu

ajutorul utilitarului Excel. În esență, programul primește la intrare unul sau mai multe fișiere text în care numără apariții ale cuvintelor, ordonate în clase. Am optat pentru genul de reprezentări grafice de tip „funcție”, clasele fiind aranjate pe axa orizontală în ordinea numerelor de cod.

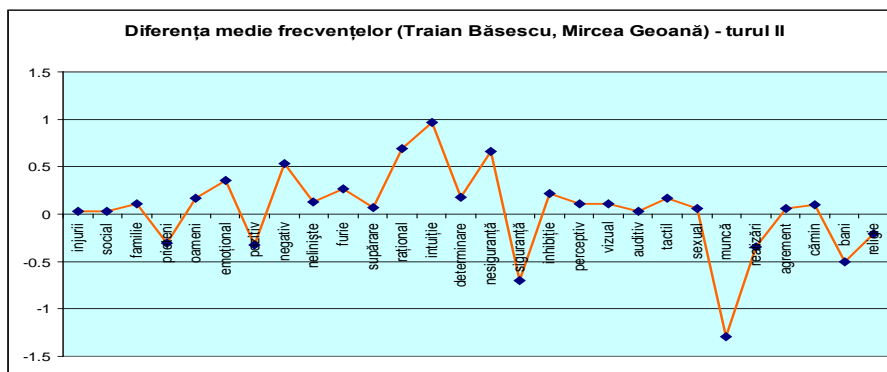
În partea a doua a evaluărilor noastre (pentru analizele comparative ale abordărilor discursive din timpul turului I și turului II) am considerat că valorile pot fi exprimate mai intuitiv prin cilindri. În interpretarea reprezentărilor grafice a editorialelor celor 3 cotidiene, dar și a discursurilor politice (atât în formă oratorică, cât și în formă dialogală) ale celor 4 candidați s-a ținut cont de pragul limită de  $\pm 0,5\%$ , considerat suficient pentru ca acea clasă care-l atinge să aibă importanță în interpretarea discursivă. Desigur, softul dă aprecieri globale în clasele respective, concluziile desprinse aici nu pleacă de la un singur text, ci reprezintă o medie pe toată colecția de texte scanate. Utilizarea reprezentărilor de tip funcție pentru punerea în evidență a unor pattern-uri discursive ale subiecților ce reprezintă, în cele din urmă, ferestre către lumile lor interioare (emoționale și cognitive), interpretarea realizându-se prin compararea datelor statistice furnizate de program, în trei moduri, pe care le-am denumit *unu-la-unu*, *unu-la-doi* și *unu-la-trei*, susținute prin formule matematice<sup>13</sup>;

Utilizarea instrumentarului informatic în aprecierea discursului politic (prelucrare, prezentare grafică și interpretare) ne-a ajutat să conturăm câteva trăsături distinctive care au dat o viziune nouă asupra tabloului discursiv al candidaților la președinție și al electoratului românesc de la sfârșitul anului 2009. Din reprezentarea grafică, luăm spre exemplificare turul II de campanie, reiese faptul că votanții care au decis numele noului președinte, conform diferențelor observate în graficele de tip *unu-la-unu* dintre *Traian Băsescu* și *Mircea Geoană* (graficul 1 - discursul oratoric), a fost preferat primul care s-a distins prin abordarea rațional-intuitivă (clasa **intuiție**), ton acuzator (clasa **emoțional\_negativ**), în timp ce la adversar, conform valorilor negative, se remarcă preocuparea pentru aspectele lucrative (clasele **muncă** și **bani**), abordându-le deschis (clasa **siguranță**).

<sup>13</sup> Graficele diferențelor *unu-la-unu* au fost construite folosind în fiecare clasă formula  $Diff_{x,y}^{1-1} = media(x) - media(y)$ , în care  $x$  și  $y$  sunt fie două cotidiene, fie doi prezidențiabili, iar  $media(x)$ , respectiv  $media(y)$  reprezintă mediile frecvențelor clasei respective pentru  $x$ , respectiv  $y$ . Graficele trebuie citite astfel: valorile aflate deasupra axei orizontale sunt cele care prevalează la cotidianul/candidatul  $x$  față de cotidianul/candidatul  $y$ , iar cele aflate dedesubtul axei orizontale sunt cele care au valori mai mari la cotidianul/candidatul  $y$  decât la cotidianul/candidatul  $x$ . O valoare de zero înseamnă egalitatea celor două mărimi comparate.

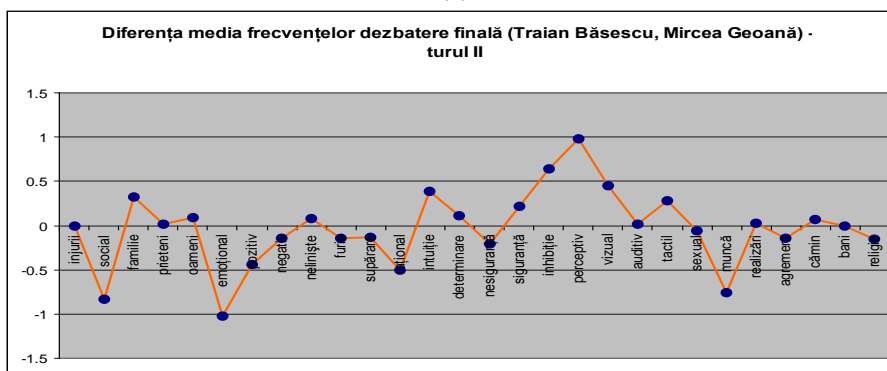
Graficele diferențelor *unu-la-doi* au fost construite folosind în fiecare clasă formula  $Diff_{x,y,z}^{1-2} = media(x) - \frac{media(y)+media(z)}{2}$ , în care  $x$ ,  $y$  și  $z$  reprezintă cele trei cotidiene ce se compară, iar  $media(x)$ ,  $media(y)$  și respectiv  $media(z)$  reprezintă mediile frecvențelor clasei respective pentru cotidienele  $x$ ,  $y$  și respectiv  $z$ . Valorile aflate deasupra axei orizontale scot în evidență clasele în care primul cotidian prevalează față de mediile celorlalte două cu care este comparat, în timp ce valorile de dedesubtul axei orizontale evidențiază clasele în care, dimpotrivă, primul cotidian se plasează sub media celorlalte două. Graficele diferențelor *unu-la-trei* au fost construite folosind în fiecare clasă formula  $Diff_{x,y,z,t}^{1-3} = media(x) - \frac{media(y)+media(z)+media(t)}{3}$ , în care  $x$ ,  $y$ ,  $z$  și  $t$  îi reprezintă pe cei patru candidați ce se compară, iar  $media(x)$ ,  $media(y)$ ,  $media(z)$  și respectiv  $media(t)$  reprezintă mediile frecvențelor clasei respective pentru candidații  $x$ ,  $y$ ,  $z$  și respectiv  $t$ . Valorile aflate deasupra axei orizontale scot în evidență clasele în care primul candidat prevalează față de mediile celorlalți trei candidați cu care este comparat, în timp ce valorile de dedesubtul axei orizontale evidențiază clasele în care, dimpotrivă, primul candidat se plasează sub media acestora.

(1)



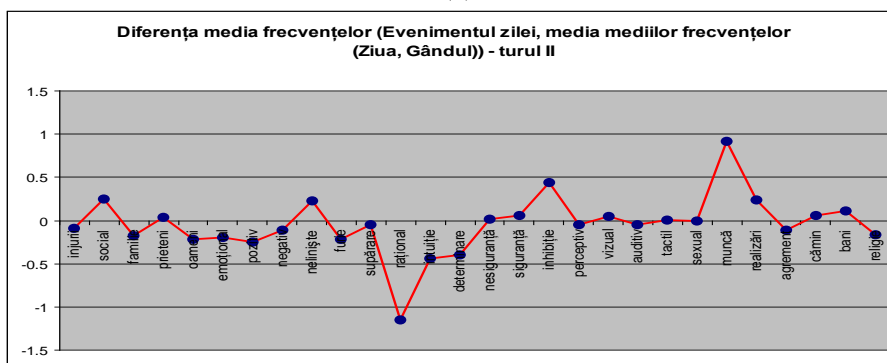
În privința comportamentului discursiv din timpul dezbaterii finale (graficul 2), *Traian Băsescu* față de *Mircea Geoană* își accentuează discursul prin cuvinte ce țin de percepții (clasa **perceptiv**), fiind mai prevăzător (clasa **inhibiție**), în timp ce adversarul său (conform valorilor negative) este mai înclinat spre afectiv (clasa **emoțional**), manifestându-se politicos (clasa **emoțional\_pozitiv**), interesul pentru aspectul lucrativ (clasa **muncă**) rămânând mult mai pregnant.

(2)



Referitor la cele trei cotidiene monitorizate în timpul campaniei prezidențiale 2009 (*Evenimentul zilei*, *Gândul* și *Ziua*), luăm spre exemplificare tot turul II de campanie. Din reprezentarea grafică, conform diferențelor observate în graficele de tip *unu-la-doi*, *Evenimentul zilei* (graficul 3) este cel mai preocupat în abordare de aspectul lucrativ (clasa **muncă**), descrierea fiind prevăzătoare (clasa **inhibiție**). Dacă privim la valorile negative, putem sublinia că descrierile sunt mai puțin raționale (clasa **rațional**), acestea fiind mai pregnante la celelalte cotidiene monitorizate.

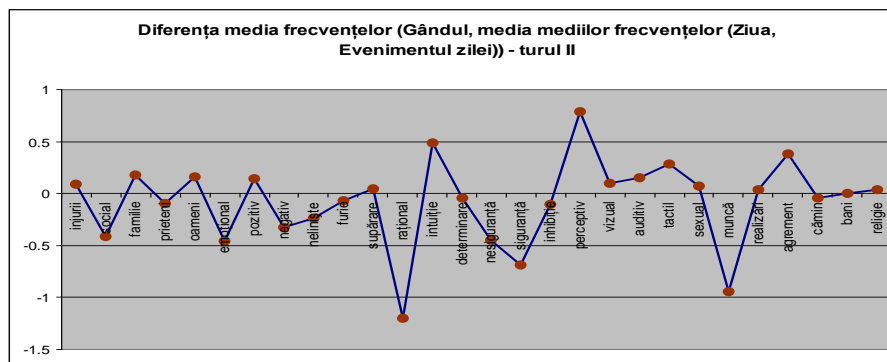
(3)





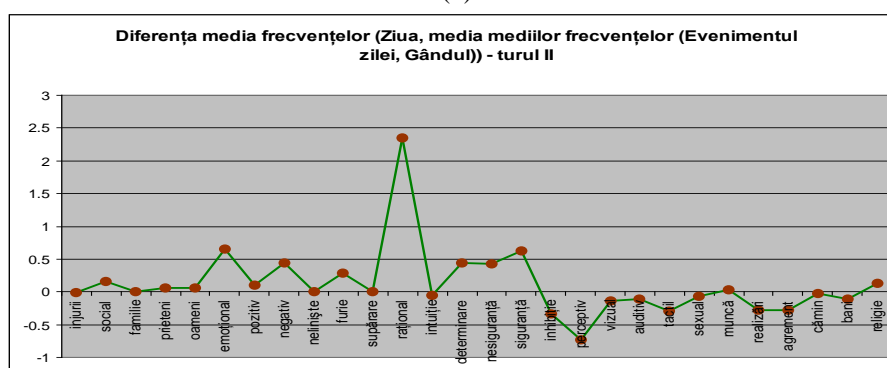
Cotidianul *Gândul* (graficul 4) se bazează cel mai mult pe percepții (clasa **perceptiv**), explicând o stare de fapt (clasa **intuiție**). Evită, conform valorilor negative, subiectele cu aspect lucrativ (clasa **muncă**) sau abordările raționale (clasa **rațional**), tonul prea afectiv (clasa **emoțional**) sau categoric (clasa **siguranță**), care nu apar la fel de pregnant ca la celelalte cotidiene monitorizate.

(4)



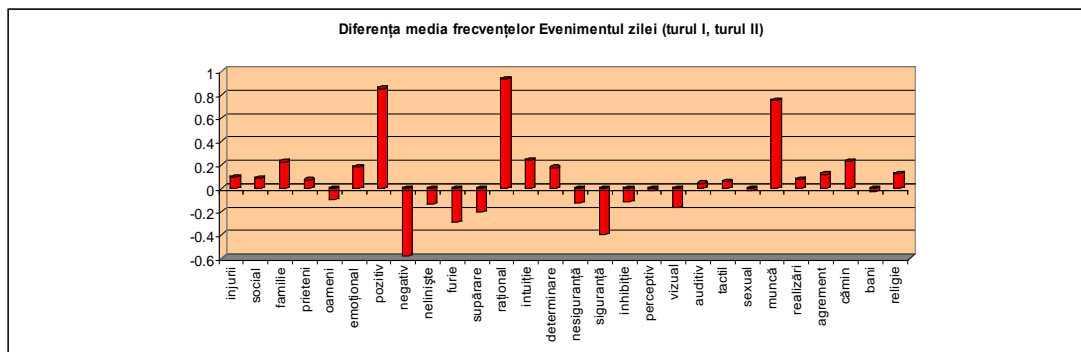
Cotidianul *Ziua* (graficul 5) se evidențiază prin capacitatea de a descrie rațional o stare de fapt (clasa **rațional**), de cele mai multe ori pe un ton categoric (clasa **siguranță**) și acuzator (clasa **emoțional\_negativ**), evitând, conform valorile negative, informarea bazată pe percepții (clasa **perceptiv**), aceasta fiind regăsită mai pregnant în editorialele celorlalte două cotidiene monitorizate.

(5)



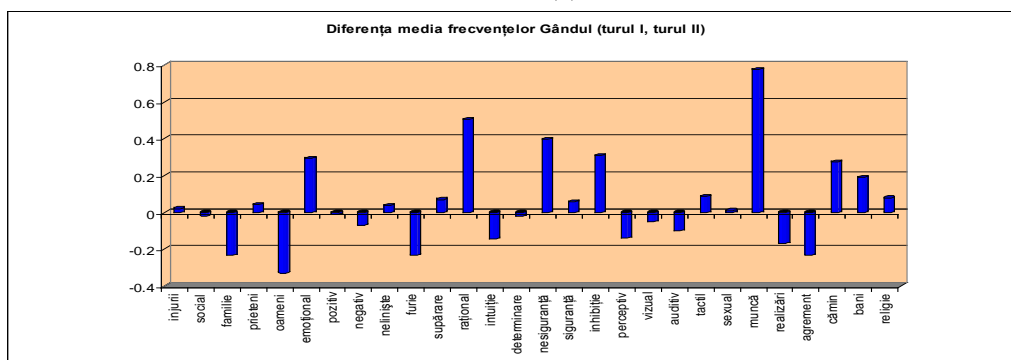
Comparând abordarea discursivă, între cele două tururi de scrutin, regăsită în editorialele celor 3 cotidiene monitorizate, constatăm din graficele de tip *unu-la-doi* că își păstrează stilul. Astfel, editorialele cotidianului *Evenimentul zilei* în turul I față de turul II de campanie (graficul 6) se înscriu în tipul de discurs rațional (clasa **rațional**), descriere populară (clasa **emoțional\_pozitiv**), un interes mai mare pentru teme din sfera lucrativă (clasa **muncă**). În turul II (conform valorilor negative) descrierea e mult mai acuzatoare (clasa **emoțional\_negativ**).

(6)



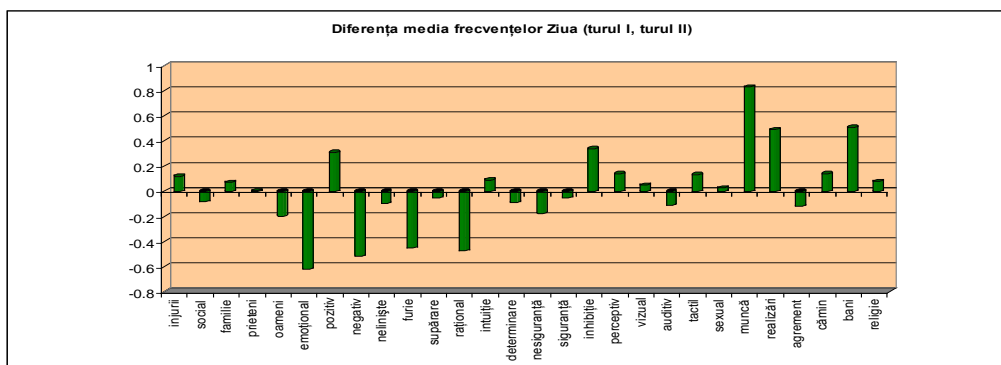
Editorialele cotidianului *Gândul* în turul I față de turul II de campanie (graficul 7) se înscriu în tipul de discurs rațional (clasa **rațional**), un interes mai mare pentru teme din sfera lucrativă (clasa **muncă**) descrise rațional (clasa **rațional**). Nu se remarcă în turul II (conform valorilor negative) schimbări radicale în tonul de abordare, dat fiind că valorile nu depășesc -0,5%, pragul minim de interpretare.

(7)



Editorialele cotidianului *Ziua* în turul I față de turul II de campanie (graficul 8) au fost mai pregnant interesate de teme din sfera economică (clasele **muncă**, **realizări** și **bani**). În turul II de campanie (conform valorilor negative) observăm o tendință mai mare pentru descrierile raționale (clasa **rațional**), pe un ton acuzator (clasa **emoțional\_negativ**).

(8)



#### 4. Concluzii

La ora actuală domeniul prelucrării limbajului natural a ajuns la o maturitate științifică și tehnologică care îl face util în activitățile derulate de cercetătorii din sfera socio-umană. Totuși, ceea ce lipsește este o interacțiune susținută și constantă a cercetătorilor care activează

în cele două domenii, care să probeze, pe de o parte, aplicabilitatea în sfera științelor sociale și umane (SSU) a tehnologiilor<sup>14</sup> create de informaticienii care se ocupă de prelucrarea textelor și, pe de altă parte, avantajele pe care cercetătorii din SSU le pot obține prin utilizarea în domeniile lor de interes a tehnologiilor lingvistice.

Pentru a-i da metodei probitatea științifică, ce permite atașarea de valori simbolice interpretative în locul unor valori numerice, mai este nevoie de încă o etapă, compararea valorilor obținute de LIWC-2007 – varianta românească existentă – cu ceea ce se știe din analize de alt tip și care au deja o recunoaștere oficială ori chiar o anumită notorietate. Dacă, de exemplu, ceea ce am aflat acum prin program (campania prezidențială 2009, supusă testului) că a reprezentat diferența cea mai notabilă între discursurile candidaților este confirmată și din alte surse, care, desigur, nu au decât posibilități de interpretare nuanțate și puțin riguroase, datele furnizate în urma acestei cercetări vor deveni suficient de relevante pentru a reliefa deja niște preferințe ale electoratului românesc către un anumit profil cognitiv definit statistic, prin mijloace riguroase, și vor putea fi preluate în viitor în arsenalul de instrumente ale campaniilor electorale.

### Mulțumiri

Teza de doctorat, care a stat la baza cercetării raportată în această lucrare a fost realizată sub îndrumarea prof. univ. dr. Laurențiu Șoitu. Softul pentru preprocesarea textelor a fost realizat de drd. Alex Moruz de la Facultatea de Informatică a Universității „Al. I. Cuza” Iași. La realizarea lexicului românesc am fost ajutată de studenții masteranzi din anul 1 (anul universitar 2009-2010) de la Masteratul de Lingvistică Computațională a aceleiași facultăți, sub îndrumarea prof. univ. dr. Dan Cristea.

### Referințe bibliografice

\* CABIN, Philippe (ed.) *La communication. État des savoirs*, Éd. Sciences Humaines, Paris, 1998

\* CRISTEA, Dan; TUFÎȘ, Dan (2002): *Resurse lingvistice românești și tehnologii informatice aplicate limbii române*, în ICHIM, O. și OLARIU, F.-T. (eds.): *Identitatea limbii și literaturii române în perspectiva globalizării*, Academia Română, Institutul de Filologie Română „A. Philippide”, Ed. Trinitas, Iași

\* EDELMAN, Murray, *Politica și utilizarea simbolurilor*, Ed. Polirom, Iași, 1999

\* GÎFU, Daniela, *Discursul presei scrise și violența simbolică. Analiza unei campanii electorale*, lucrare de doctorat susținută la Facultatea de Filosofie, Universitatea „Al. I. Cuza”, Iași

(v.

[http://www.uaic.ro/uaic/bin/download/Academic/Doctorate\\_martie\\_2010/GfuC.Daniela.pdf](http://www.uaic.ro/uaic/bin/download/Academic/Doctorate_martie_2010/GfuC.Daniela.pdf))

---

<sup>14</sup> A se vedea Dan CRISTEA, Dan TUFÎȘ (2002): *Resurse lingvistice românești și tehnologii informatice aplicate limbii române*, în O. ICHIM și F.-T. OLARIU (eds.): *Identitatea limbii și literaturii române în perspectiva globalizării*, Academia Română, Institutul de Filologie Română „A. Philippide”, Ed. Trinitas, Iași

- \* PENNEBAKER, James W., FRANCIS, Martha E., BOOTH, Roger J., *Linguistic Inquiry and Word Count – LIWC2001*, Erlbaum Publishers, Mahwah, NJ, 2001 și pe [www.erlbaum.com](http://www.erlbaum.com)
- \* PERLMUTTER, David D., *The Manship School guide to political communication*, Baton Rouge, Louisiana State University Press, 1999
- \* SINESCU, Călin, *Comunicare politică*, Ed. Universitară, București, 2007
- \* ȘOITU, Laurențiu, *Philosophy of communication*, Ed. Institutul European, Iași, 2001
- \* WOLTON, Dominique, *Les contradictions de la communication politique*, CNRS Éditions, Paris, 1995