

Investește în oameni !

FONDUL SOCIAL EUROPEAN

Program Operațional Sectorial Dezvoltarea Resurselor Umane 2007-2013

Axa prioritară 1 Educația și formarea profesională în sprijinul creșterii economice și dezvoltării societății bazate pe cunoaștere;

Domeniul major de intervenție 1.5 Programe doctorale și post-doctorale în sprijinul cercetării;

Titlul proiectului: „**Rețea transnațională de management integrat al cercetării postdoctorale în domeniul Comunicarea științei. Construcție instituțională (școală postdoctorală) și program de burse (CommScie)**”

Numărul de identificare al contractului: POSDRU/89/1.5/S/63663

Beneficiar: Universitatea “Alexandru Ioan Cuza” Iași

RAPORT DE CERCETARE**Cod raport: ANUL I-B (luna 1-12)** (Selectați)**Perioada de raportare: octombrie 2010 - septembrie 2011** (Selectați)**1. DATE PERSONALE ALE CERCETĂTORULUI POST-DOCTORAL:**

Nume:	GÎFU
Prenume:	DANIELA
Telefon:	0742050673
E-Mail:	gifudaniela@yahoo.com ; daniela.gifu@info.uaic.ro

2. DATE PERSONALE ALE TUTORELUI CERCETĂTORULUI POST-DOCTORAL:

Nume:	CRISTEA
Prenume:	DAN
Telefon:	0232-201.542
E-Mail:	dcristea@info.uaic.ro

3. INSTITUȚIA GAZDĂ A CERCETĂTORULUI:

Denumire Institutie:	Universitatea Alexandru Ioan Cuza, Iasi	(Selectați)
Domeniul fundamental de cercetare :	D 4 – Științe ale omului și ale societății	(Selectați)
Facultate/ Department:	Informatică/Cercetare Informatică & Psihologie și Științe ale Educației/Cercetare Interdisciplinară în științe umaniste	

4. TITLUL PROIECTULUI DE CERCETARE:

**Utilizarea tehnologiilor limbajului natural
în măsurarea reacțiilor mass-mediei la evoluțiile de pe scena politică**

5. OBIECTIVELE PROIECTULUI (PENTRU PERIOADA DE RAPORTARE):

Obiective prevăzute	Obiective realizate	Gradul de realizare (total/ parțial/ nerealizat)	Descriere / observații:
---------------------	---------------------	---	-------------------------

1	Completarea lexicului prin cuvinte relevate de un studiu statistic pe corpusuri ale discursului politic românesc din presa scrisă în diferite contexte socio-politice.	- soluționarea sursei de erori raportate anterior în statistici; - rafinarea analizei încărcăturii emoționale și cognitive într-un discurs politic. Au fost realizate 3 interfețe diferite (AnaDiP, PLP și PEDANT). În prezent, se lucrează cu 29 de clase semantice.	total	clarificarea importanței diferite a lexicalilor.
2	Sofisticarea softwareului în vederea semnării inclusiv a cuvintelor din interiorul claselor care au dus la valorile calculate.	- adăugarea în lexicul programului a două forme de intrări lexicale: de tip <word-root>(*)-<sem-list> și <lemma><POS><sem/list>	total	cuplarea DEX-online la lexicul programului și asignarea cuvintelor la cele 29 de clase semantice.
3	Utilizarea reprezentărilor de tip „funcție”, „coloană”, „diagramă circulară”, „diagramă cu domenii” pentru punerea în evidență a unor patternuri discursive ale subiecților monitorizați.	- interpretarea graficelor prin compararea datelor statistice furnizate de program; - realizarea unei biblioteci cu formule matematice necesare analizelor comparative sau ratei de apariție cuvânt / sintagmă;	total	compararea rezultatelor obținute cu formulele matematice existente în soft cu altele obținute prin metode statistice.

6. ACTIVITĂȚILE PROIECTULUI (PENTRU PERIOADA DE RAPORTARE):

	Activități prevăzute	Activități realizate	Gradul de realizare	Descriere/observații:
1	Documentare cu privire la existența, importanța și relevanța altor analizori statistici.	- testarea altor resurse lingvistice: WordNet, AnaMorph, RACAI, T-LAB etc. - validarea softwareului prin compararea rezultatelor obținute cu alte programe de acest tip.	total	Crearea unei legături între proiectele informatice de același profil

2	Completarea și optimizarea lexicului programului; asignarea cuvintelor / lemelor la clasele semantice.	- cuplarea DEX-online și evaluarea lexicurilor pe cele 29 de clase semantice; - îmbunătățirea claselor în funcție de specificul limbajului politic românesc.	total	Rafinarea rezultatelor oferite de noile interfețe la care s-a lucrat în identificarea tonalităților discursiv-politice din presa scrisă
3	Promovarea activității de cercetare la diverse manifestări științifice naționale și internaționale cu caracter formal sau informal.	- diseminare în țară: Târgu Mureș, Cluj-Napoca, Iași, Suceava. - diseminare în străinătate: Croația, Ungaria, Grecia, SUA, Italia.	total	Cunoașterea parcursului cercetării noastre în domeniul NLP (Natural language processing) atât în țară, cât și în străinătate cu posibilitatea încheierii unor parteneriate viitoare.

7. REZULTATE LIVRATE (ÎN PERIOADA DE RAPORTARE) :

7.1. Participări la conferințe naționale:

Titlul conferinței	Data și locul desfășurării	Organizator	Titlul lucrării prezentate	Calitate (autor, coautor etc.)
<i>Comunicare, context, interdisciplinaritate</i>	19-20 nov. 2010	Universitatea „Petru Maior” Târgu-Mureș	<i>Utilizarea tehnologiilor de prelucrare a limbajului natural în context electoral: Metoda LIWC-2007</i>	autor
<i>Dezvoltare comunitară și incluziune socială în perspectivă socio-economică</i>	10-11 dec. 2010	Universitatea „Alexandru Ioan Cuza” Iași	<i>Discursul electoral – formă de influențare psihosocială. Tehnici computaționale de prelucrare lingvistică</i>	autor
<i>1 to 1. Networking meeting research – industry, Bring IT on</i>	10.12.10	Universitatea „Alexandru Ioan Cuza” Iași	1. <i>Analiza discursului politic</i> 2. <i>Sistem de flexionare pentru limba română</i>	1. autor 2. co-autor
<i>Prelegere în cadru informal</i>	04.01.11	Universitatea Populară Cluj-Napoca	<i>Tehnologii de prelucrare a limbajului politic - AnaDiP</i>	autor
Simpozion: <i>Limbajul politic, încotro?</i> în cadru informal	18 feb. 2011	Centrul de Dezbateri Socio-politice „Emil Dandea” (Palatul Apollo) Târgu-Mureș	<i>Radiografierea discursului electoral. Metoda AnaDiP-2010.</i>	autor

<i>Democratizare, descentralizare și asistență socială. Prezent și viitor</i>	5-7 mai 2011	Universitatea „Ștefan cel Mare” Suceava	<i>Utilizarea tehnologiilor de prelucrare a limbajului natural în contextul schimbărilor politice</i>	autor
<i>Natural Language Processing Goes Industrial</i>	28 aug.-4 sept. 2011	Universitatea „Babeș-Bolyai” Cluj-Napoca	<i>Using natural language technology to measure mass media reactions in the election context</i>	co-autor

7.2. Participări la conferințe internaționale

Titlul conferinței	Data și locul desfășurării	Organizator	Tipul prezentării invitat/oral/poster	Titlul lucrării, autorii, afilierea
<i>Information Society and Globalisation: Transformation of Politics</i>	22 oct. 2010 University of Dubrovnik	University of Dubrovnik	invitat	<i>Using Natural Language Processing Technologies: Method LIWC-2007,</i> Daniela Gifu și Dorin Popa
<i>ICT (Information and Communication Technologies) Proposers' Day 2011</i>	19-20 mai 2011, Hungexpo, Budapest	Hungarian EU Presidency	invitat	<i>Computational techniques applied to the analysis of the political language:</i> AnaDiP, Daniela Gifu
<i>The 6th International Conference on Future Information Technology - FutureTech2011 - Greece</i>	28-30 iun. 2011, Loutraki, Greece	Seoul National University of Technology, Korea	oral	<i>Computational Techniques in Political Language Processing:</i> AnaDiP-2011, Daniela Gifu and Dan Cristea
<i>The 2011 World Congress in Computer Science, Computer Engineering and Applied Computing (WORLDCOMP'11) - To the 2011 International Conference on Artificial Intelligence (ICAI'11)</i>	18-21 iul. 2011, Las Vegas, USA	University of Georgia, Georgia, USA	oral	<i>The Analysis of print media discourse in the election context,</i> Daniela Gifu
<i>The International Workshop on Intelligent Information Systems IIS'2011</i>	13-14 sept. 2011, Chișinău, Rep. Moldova	Institute of Mathematics and Computer Science, Chișinău	oral	<i>A technique for political language processing,</i> Daniela Gifu

<i>Colloque international Les Sciences sociales et leurs publics. Engagements distanciations</i> et	22-24 sept. 2011, Iași	Universitatea „Alexandru Ioan Cuza” Iași	membru în Comitetul de organizare
---	---------------------------	---	-----------------------------------

7.3. Cărți publicate

Titlul cărții	Editură	An apariție	ISBN/ISSN	Număr de pagini	Calitate (autor, coautor, coordonator etc.)
<i>Violența simbolică în discursul electoral</i>	Casa Cărții de Știință, Cluj-Napoca	2011	ISBN	398	autor
<i>Dezvoltare comunitară și incluziune socială în perspectivă socio-economică</i>	Ed. Universității „Alexandru Ioan Cuza”, Iași	2011	ISSN	sub tipar	co-editor

7.4. Articole/Lucrări publicate:

7.4.1. Lucrări publicate în reviste cotate ISI

Autor(i)	Titlul lucrării	Revista	Data apariției	ISBN/ISSN	Pagini

7.4.2. Lucrări publicate în reviste indexate în BDI

Autor(i)	Titlul lucrării	Revista	Data apariției	ISBN/ISSN	Pagini
Daniela Gîfu	<i>Emotional load identification in the Romanian electoral discourse. The technique AnaDiP</i>	Cultus – The Journal of Intercultural Mediation and Communication, Italy	sub tipar	ISSN	-

7.4.3. Lucrări publicate în BDI

Autor(i)	Titlul lucrării	BDI	Data apariției	Adresa web
Daniela Gîfu, Dan Cristea	<i>Computational Techniques applied to the analysis of the political language: AnaDiP</i>	Springer	Iunie, 2011	http://www.ftrai.org/futuretech2011/pop_accepted.htm#music

Daniela Gîfu	<i>The Analysis of print media discourse in the election context</i>	Springer	Iulie, 2011	http://www.ucmss.com/cr/maian/papersNew/papersAll/ICA3324.pdf
Daniela Gîfu, Dan Cristea	<i>Using natural language technology to measure mass media reactions in the election context</i>	Springer	August, 2011	http://www.easychair.org/conferences/submission.cgi?a=a10c89b0ad82;track=44433;submission=806656
Daniela Gîfu	<i>A technique for processing of political language</i>	Mathematical Reviews and Zentralblatt MATH	Septembrie, 2011	http://www.math.md/iis2011/II_S2011%20Program%20modified%20September%202009.pdf

7.4.4. Lucrări publicate în reviste românești recunoscute de CNCSIS

Autor(i)	Titlul lucrării	Revista	Data apariției	ISBN/ISSN	Pagini	Indexarea revistei (B, B+, C)
Daniela Gîfu	<i>Political language processing. Method AnaDiP</i>	Annals of „Ștefan cel Mare” University of Suceava	(sub tipar) 2011	ISSN	-	B+

7.4.5. Lucrări publicate în volumele unor conferințe internaționale

Autor(i)	Titlul lucrării	Volumul	Editor coordonator	Data apariției	ISBN/ISSN	Pagini
Daniela Gîfu, Dan Cristea	<i>Computational Techniques applied to the analysis of the political language: AnaDiP</i>	FutureTech 2011, Part II, CCIS 185	J.J. Park, L.T. Yang, and C. Lee	Iunie, 2011	LNCS	188–195
Daniela Gîfu	<i>The Analysis of print media discourse in the election context</i>	ICAI 2011, vol. II, CSREA Press, USA,	Hamid R. Arabnia, David de la Fuente Elena B. Kozerenko, Jose A. Olivas	Iulie, 2011	LNCS	580–586
Daniela Gîfu	<i>A technique for processing of political language</i>	Proceedings IIS, International Workshop on Intelligent Information Systems	Institute of Mathematics and Computer Science, Academy of Sciences of Moldova	Septembrie, 2011	MATH	175-178

7.4.6. Lucrări publicate în volumele unor conferințe naționale / cu participare internațională

Autor(i)	Titlul lucrării	Volumul	Coord. volum	Editura	Data apariției	ISBN/ISSN	Pagini
Daniela Gîfu	<i>Utilizarea tehnologiilor de prelucrare a limbajului natural în context electoral</i>	Comunicare, context, interdisciplinabilitate	Iulian Boldea	Ed. Univ. „Petru Maior” Târgu Mureș	2010	ISSN	87-98
Daniela Gîfu	<i>Discursul electoral – formă de influențare psihosocială</i>	Comunicare, context, interdisciplinabilitate	Daniela Gîfu și Mihai Dinu Gheorghiu	Ed. Univ. „Alexandru Ioan Cuza” Iași	(sub tipar) 2011	ISSN	-

7.4.7. Alte publicații (de popularizare/comunicare a științei)

Autor(i)	Titlul lucrării	Revista/cotidian	Data apariției	ISBN/ISSN	Pagini
Valentin Huțanu	<i>La Cuza s-a inventat detectorul de minciuni al politicianilor</i>	Bună ziua, Iași	10.12.10	-	7
Carmen Baraboi	<i>Proiecte inovatoare prezentate la UAIC</i>	Evenimentul, Iași	11.12.10	-	4
A.Z.	<i>Discursul politic, încotro?</i>	Zi de zi, Târgu-Mureș	18-20 februarie 2010	-	1
-	<i>În an preelectoral DEZBATERE DESPRE DISCURSUL POLITIC</i>	Cuvântul liber, Târgu-Mureș	17.02.10	-	3
-	<i>Simpozion: Limbajul politic – încotro?</i>	Cuvântul liber, Târgu-Mureș	18.02.10	-	4
-	<i>Prelegere despre discursul politic</i>	Făclia, Cluj-Napoca	04.01.11	-	9

7.5. Brevete depuse spre omologare:

7.6. Brevete omologate:

7.7. Cereri de finanțare/aplicații depuse:

8. STAGII DE CERCETARE ȘI DOCUMENTARE ÎN STRĂINĂTATE :

Locul (instituția: universitatea, institutul de cercetări etc.)	Perioada	Numărul de săptămâni

9. STAGII DE CERCETARE ȘI DOCUMENTARE ÎN ȚARĂ:

Locul (instituția: universitatea, institutul de cercetări etc.)	Perioada	Numărul de săptămâni
Universitatea „Babeș-Bolyai”, Cluj-Napoca	28 aug.-4 sept. 2011	1

10. PARTICIPĂRI ÎN ECHIPE DE CERCETARE ALE ALTOR PROIECTE:

Proiectul (denumire și cod proiect/nr. contract etc.)	Programul	Valoarea proiectului	Calitatea (cercetător, asist.manager etc.)	Perioada

11. Publicațiile sau rezultatele apărute și raportate în urma cercetării finanțate prin proiectul POSDRU/89/1.5/S/63663 au menționat numele finanțatorului și numărul de contract:

DA (Selectați)

**PRIN ACEASTA SE CERTIFICA LEGALITATEA SI CORECTITUDINEA
DATELOR CUPRINSE IN PREZENTUL RAPORT DE ACTIVITATE și în ANEXA 1 .**

Data completării:
30/09/2011 (Selectați)

Nume, prenume
Gifu, Daniela
Semnătura

Secțiune destinată tutorelui cercetătorului post-doctoral:

**AVIZUL TUTORELUI PENTRU
CONTINUAREA ACTIVITĂȚII DE CERCETARE:**

([Selectați](#))

Pe parcursul primului an de cercetare postdoctorală, doamna dr. Daniela Gîfu s-a implicat cu toată seriozitatea în activitatea științifică din cadrul departamentului de cercetare informatică, realizând mult peste obiectivele propuse inițial. Munca doamnei Gîfu este deja cunoscută atât în țară, cât și în străinătate, datorită participării sale la multe manifestări științifice de profil NLP (procesarea limbajului natural). De asemenea, trebuie să spunem că rezultatele obținute până în prezent au stat la baza unor lucrări științifice importante, deschizând un drum important în domeniul procesării limbajului natural cu aplicabilitate în Filosofie, Jurnalism, Sociologie s.a.m.d.

Sunt întru totul de acord să-și continue programul de cercetare din cadrul proiectului POSDRU/89/1.5/S/63663.

**AVIZUL TUTORELUI PENTRU VALIDAREA
FINALĂ A ACTIVITĂȚII DE CERCETARE:**

([Selectați](#))

(se bifează doar pentru ultimul raport de cercetare al bursei):

Data avizării:
30/09/2011 ([Selectați](#))

Nume, prenume
CRISTEA, Dan
Semnătura

RAPORT DE CERCETARE (rezumat extins)

1. Stadiul cunoașterii și cercetărilor în domeniu, la nivel internațional și național. Delimitări terminologice și conceptuale. Bibliografia relevantă în domeniul de cercetare. (max. 2 pag.)

Domeniul NLP (*Natural Language Processing*) a devenit un exercițiu eficient de interacțiune în probarea aplicabilității în sfera științelor umaniste și sociale (SSU) a tehnologiilor lingvistice dezvoltate de informaticienii lingviști (specialiști ai domeniului lingvistică computațională). Încă din anii 50'-60' s-a simțit nevoia informatizării limbilor naturale, modelării și construirii bazelor de date lingvistice și extralingvistice pentru prelucrarea mecanică a limbilor naturale.

Plecând de la un lexic insuficient pentru tipul de analiză propusă, s-a trecut la etapa de completare a lexicului prin cuvinte relevate de un studiu statistic pe corpusuri ale discursului politic românesc din presa scrisă în diferite contexte socio-politice. Etapa ce a urmat, a constat în colectarea și stocarea unui nou corpus lingvistic cu specific politic. Prin corpus lingvistic, pe scurt *corpus*, se înțelege o colecție mare de date lingvistice/dovezi asupra limbii. În general, poate să includă: conversații de zi cu zi, știri radio, scrieri publicate și chiar scrierile dedicate copiilor. Noi ne-am oprit doar la texte politice în format electronic, cunoscut fiind că presa scrisă dispune și de pagini web de promovare.

Scopul proiectului nostru este acela de a prezenta și interpreta contextul socio-politic, în diverse etape istorice, totodată, cu ajutorul tehnicilor computaționale de prelucrare lingvistică a discursurilor politice din presa scrisă, tipărită sau online. Concepția care stă la baza acestei metode este aceea că maniera în care indivizii vorbesc și scriu, în încercarea de a-și impune imaginea personală valorizată, care le poate asigura credibilitate, reprezintă o legătură între dimensiunea lor emoțională și cea cognitivă. Prin accentuarea componentei afective la nivel de discurs, alocutorii (cititorii amatori sau profesionalizați) se identifică cu locutorul (actorul politic, jurnalistul), care devine vocea legitimă a idealurilor comune.

Efortul depus până în prezent, de înțelegere și interpretare a limbajului politic, s-a concretizat în realizarea unei noi resurse lingvistice PEDANT (*Political and Emotional Discourse Analysis Tool*) (în trei etape diferite constituite de cele trei interfețe la care s-a lucrat în cadrul Facultății de Informatică: AnaDiP, *Analiza Discursului Politic*, PLP, *Political Language Processing* și PEDANT). Noua interfață evidențiază cu claritate și obiectivitate influența discursului politic asupra receptorilor săi, cunoscut fiind că discursivitatea politică instituie o realitate posibilă și alternativă. Acest software a avut ca surse de inspirație alte resurse lingvistice, ca:

- LIWC (Linguistic Inquiry and Word Count), dezvoltată în colaborare de Universitatea din Texas at Austin și Universitatea din Auckland), v. www.liwc.net.
- WordNet – Universitatea Princeton și versiunea românească realizată în cadrul proiectului Balkanet), precum și tehnologia NooJ (<http://www.nooj4nlp.net/pages/nooj.html>).
- AnaMorph – versiunea românească de flexionare a limbii române realizată în cadrul Facultății de Informatică, Universitatea „Alexandru Ioan Cuza” Iași.
- RACAI - (*Romanian Academy Center for Artificial Intelligence*), resursă lingvistică bilingvă: română și engleză. Colaborare între Institutul de Cercetări pentru Inteligență Artificială și Academia Română.

– T-LAB, www.tlab.it/default.php, soft italian în vederea comparării și validării rezultatelor obținute cu AnaDiP.

Rezultatele obținute, atât cantitativ (valori procentuale corespunzătoare fiecărei clase semantice din cele 29 de care dispune programul în acest moment), cât și calitativ (patru tipuri de reprezentări grafice de tip „funcție”, „coloană”, „diagramă circulară”, „diagramă cu domenii” care pun în evidență unele patternuri discursive ale subiecților monitorizați).

Metoda computațională de analiză a conținutului politic vine în sprijinul cercetătorilor-specialiști în științe social-politice, analiști politici, jurnaliști, staff-uri electorale, fiind de ajutor, cu preponderență, în prospecțiunile sociale atașate campaniilor de comunicare politică (cu preponderență în context (pre)electoral), în vederea măsurării reacțiilor mass-mediei la evoluțiile de pe scena politică.

Prezentăm și o scurtă bibliografie care a stat la baza cercetării noastre:

1. G. Edward Barton, Robert C. Berwick at al. (1987), „Computational complexity and natural language”, The MIT Press, Cambridge, Massachusetts London, England.
2. J.-P. Bronckart, D. Bain, B. Schneuwly, C. Davaud et A. Pasquier (1985), “Le fonctionnement des discours. Un modèle psychologique et une méthode d’analyse”, Delachaux et Niestlé.
3. Dan Cristea, Dan Tufiș (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.
4. Barbara Fox (1987), „Discourse structure and anaphora. Written and conversational English”, Cambridge University Press
5. Daniela Gîfu (2010), “Discursul presei scrise și violența simbolică. Analiza unei campanii electorale”, lucrare de doctorat susținută la Facultatea de Filosofie și Științe Social-Politice, Universitatea „Alexandru Ioan Cuza”, Iași.
6. Daniela Gîfu, Dan Cristea (2011), “Computational Techniques in Political Language Processing: AnaDiP-2011” in J.J. Park, L.T. Yang, and C. Lee (Eds.): “FutureTech 2011”, Part II, CCIS 185, pp. 188–195, Springer-Verlag Berlin Heidelberg.
7. Christiane Fellbaum (ed.) (2001), “WordNet, An Electronic Lexical Database”, The MIT Press.
8. J. T. Hancock, L. E. Curry, S. Goorha and M. Woodworth: “On lying and being lied to: A linguistic analysis of deception in computer-mediated communication.” *Discourse Process* 45 (1).
9. Michael Lesk (1986), “Automatic sense disambiguation” How to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC Conference*, New York, Association for Computing Machinery.
10. Christopher D. Manning, Hinrich Schütze (2003) “Foundations of Statistical Natural Language Processing”, London, England, The MIT Press Cambridge
11. R. Mihalcea and C. Strapparava (2009) “The Lie Detector: Exploration in the Automatic Recognition of Deceptive Language”. *Proceedings of the Association for Computational Linguistics (ACL)*.
12. James W. Pennebaker, Martha E. Francis, Roger J. Booth (2001), “Linguistic Inquiry and Word Count – LIWC2001”, Mahwah, NJ, Erlbaum Publishers.
13. David D. Perlmutter (1999), “The Manship School guide to political communication”, Baton Rouge, Louisiana State University Press.
14. Stuart L. Pook, Jason Catlett (1988), “Making sense out of searching”. In *Information Online 88*, Sydney. The Information Science Section of the Librarian Association of Australia.
15. Petr Sojka, Karel Pala et al. (2004), „Proceedings of the Second International WordNet Conference GWC 2004”, Masaryk University, Brno, Czech Republic.
16. Laurențiu Șoitu (2001), *Philosophy of communication*, Institutul European, Iași.
17. Alain Touraine (1992), “Critique de la modernité”, Fayard, Paris.
18. Dan Tufiș (2001), „Promovarea limbii române în SI – SC”, în Florin Gh. Filip (ed.), *Societatea Informațională - Societatea Cunoașterii, concepte, soluții și strategii pentru România*, pp. 131-142, Academia Română, Ed. Expert, București.
19. Dominique Wolton (1995), “Les contradictions de la communication politique”, CNRS, Paris.

2. Obiectivele generale ale proiectului. (max. 1/2 pag.).

Insuficiența instrumentarului informatic în interpretarea limbajului politic la momentul septembrie 2010 a atras după sine și creionarea unor noi direcții de cercetare, devenite obiectivele generale ale acestui proiect.

1. Cunoașterea altor resurse lingvistice care au drept scop analiza de conținut în vederea validării programului nostru.
2. Completarea lexicului prin cuvinte relevate de un studiu statistic pe corpusuri ale discursului politic românesc și o sofisticare a programului în vederea semnalării inclusiv a cuvintelor din interiorul claselor care au dus la valorile calculate. S-a urmărit rafinarea analizei încărcăturii emoționale și cognitive într-un discurs politic. Au fost realizate 3 interfețe diferite (AnaDiP, PLP și PEDANT), cu 29 de clase semantice.
- 3.. Utilizarea reprezentărilor de tip „funcție”, „coloană”, „diagramă circulară”, „diagramă cu domenii” pentru punerea în evidență a unor patternuri discursive ale subiecților monitorizați.
4. Includerea în cadrul programului PEDANT a două forme de intrări lexicale: de tip <word-root>(*)-<sem-list> și <lemma><POS><sem/list>, unde <sem-list> este lista celor 29 de clase semantice. Noul soft dispune de două variante de lexic: *lexicul inițial îmbunătățit* - în prezent include peste 6.000 cuvinte/rădăcini și *lexicul cu leme* – 76.760 cuvinte formă de bază și 63.344 cuvinte formă derivată.
5. Includerea unei biblioteci cu formule matematice necesare diverselor analize comparative: discursurile politice ale aceluiași candidat/ziar în două perioade diferite; discursurile politice ale doi sau mai multor candidați/ziare într-o perioadă dată etc.

3. Metodologia utilizată (max. 1 pag.).

În elaborarea concluziilor preliminare privind desfășurarea unui proces discursiv (s-a pornit de la contextul electoral din 2009) au fost culese, stocate și centralizate electronic texte politice, editorialele unor publicații naționale cu profil comun, precum și discursurile politice (în ambele forme: oratorică și dialogică) a celor patru prezidențiabili din acel an în România. Acestea au stat la baza alegerii claselor semantice cărora le-au fost construite lexici adecvate. Acele clase răspund foarte bine intenției de identificare a tonalităților discursive cu efect persuasiv, necesare legitimării unei anumite voci politice.

Prezentăm pe scurt pașii metodologici urmați în realizarea programului PEDANT. Cea mai dificilă etapă a fost aceea de construire a lexicului.

- Mai întâi s-au tradus termenii englezești de la LIWC, în ordine alfabetică, regăsiți în cele 28 de clase reținute în interfața inițială din cele 64 de care dispunea softul american. În această etapă s-a decis, pentru fiecare cuvânt ori rădăcină inclusă, care trebuie să fie familia de rădăcini corespunzătoare semnificației sale, 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;
- Apoi au fost sortate, în ordine alfabetică, cuvintele/rădăcinile pentru fiecare categorie în parte;
- 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);
- Toate dicționarele asignate claselor au fost puse împreună, lista astfel obținută s-a sortat din

nou alfabetice și s-au eliminat aparițiile multiple (prin concatenarea listelor de clase, cu lăsarea a singură dată a unui număr de clasă în lista unui cuvânt/unei rădăcini);

- În cele din urmă, pornind de la îmbinări de cuvinte, au fost introduse și o parte din cele mai uzitate expresii;

- Etapa de dezvoltare a lexicului a necesitat și cuplarea la DEX-online. Ne aflăm într-un studiu de îmbogățire automată, pornind de la textele politice românești din prezent (campanii de comunicare politică).

Softul PEDANT (al cărui parametru este calculul de frecvență - analiza cantitativă), rezultatele fiind apoi interpretate automat (analiza calitativă) pe baza reprezentărilor grafice oferite de program. În esență, programul primește la intrare unul sau mai multe fișiere text în care numără apariții ale cuvintelor, ordonate în clase (29, în prezent): <class id="1" name="Injurii"/>, <class id="2" name="Social"/>, <class id="3" name="Familie" parent="2"/>, <class id="4" name="Prieteni" parent="2"/>, <class id="5" name="Oameni" parent="2"/>, <class id="6" name="Emoțional"/>, <class id="7" name="Emoțional_pozitiv" parent="6"/>, <class id="8" name="Emoțional_negativ" parent="6"/>, <class id="9" name="Anxietate" parent="8"/>, <class id="10" name="Furie" parent="8"/>, <class id="11" name="Supărare" parent="8"/>, <class id="12" name="Rațional"/>, <class id="13" name="Intuiție" parent="12"/>, <class id="14" name="Determinare" parent="12"/>, <class id="15" name="Nesiguranță" parent="12"/>, <class id="16" name="Siguranță" parent="12"/>, <class id="17" name="Inhibiție" parent="12"/>, <class id="18" name="Perceptiv"/>, <class id="20" name="Auditiv" parent="18"/>, <class id="19" name="Vizual" parent="18"/>, <class id="21" name="Tactil" parent="18"/>, <class id="22" name="Sexual"/>, <class id="23" name="Muncă"/>, <class id="24" name="Realizări"/>, <class id="25" name="Nerealizări"/>, <class id="26" name="Agreement"/>, <class id="27" name="Cămin"/>, <class id="28" name="Financiar"/>, <class id="29" name="Religie"/>.

4. Rezultatele obținute și diseminarea acestora (impactul, relevanța și aplicabilitatea rezultatelor) (max. 1 pag).

În acest an de cercetare postdoctorală ne-am oprit la analiza lexicală a discursului politic scris, luând în considerare doar informația scrisă, nu și informația transmisă vocal. Așa a apărut ideea implementării unei interfețe românești, realizată de masteranda Mădălina Spătaru sub îndrumarea prezentului cercetător postdoc. Desigur, vizualizarea rezultatelor analizei necesită existența unui corpus textual considerabil, dat fiind că lexicul se află într-o permanentă etapă de îmbunătățire (segmentat în mai multe fișiere de tip .doc sau .txt). Sunt, apoi, selectate pe rând toate acele fișiere care fac obiectul analizei propuse. Există diferite modalități de afișare a rezultatelor procesării, inclusiv opțiunea de evidențiere a cuvintelor în text ce determină tipul discursului (public, politic, literar, medical etc.). Intenția noastră, anul II de cercetare postdoctorală, este aceea de realizare a unui instrument capabil să detecteze și să analizeze orice tip de discurs.

În primă instanță va fi detectat tipul discursului pentru ca ulterior să fie luate în considerare doar informațiile relevante contextului emiterii mesajului. Se consideră că unui discurs îi pot fi atribuite mai multe categorii (clase) semantice care au grade diferite de relevanță. Rata de apariție, intensitatea emoțională, forța explicativă etc. Se creează premisa stabilirii unor concluzii ce pot sprijini definirea tipului de mesaj emis.

Pe baza acestor categorii semantice (29, în prezent), sunt realizate statistici și comparații cu alte discursuri din același domeniu (politic, în cazul nostru). S-a urmărit crearea unei legături între proiectele informatice în vederea rafinării rezultatelor oferite de noua metodă computațională PEDANT în identificarea tonalităților discursiv-politice (cu accent pe aspectele emoționale și raționale), necesare cercetării prezente. Aceasta a necesitat cunoașterea unei părți a tezaurului lexical (s-a lucrat și în echipa celor implicați în proiectul E-DTLR, care dispune de o imensă bază de date) ce a fost utilizată

pe parcursul anului octombrie 2010 – septembrie 2011.

Existența celor trei etape de implementare a softului actual a însemnat și îmbunătățirea euristiciilor de utilizare a programului, în vederea creșterii acurateții datelor rezultate în urma procesării textelor politice, inclusiv cele din sec. XIX.

De altfel, fiecare pas întreprins până în prezent a fost promovat în numeroasele lucrări publicate în diverse reviste științifice, *Proceedings*-urile conferințelor, volume (în calitate de autor, co-autor sau colective) atât în țară (Târgu-Mureș, Cluj-Napoca, Suceava, Iași), cât și în străinătate (Croatia, Ungaria, Grecia, SUA, Republica Moldova, Italia), în prelegerile cu caracter științific susținute în cadru formal (Universități de prestigiu din țară, Europa și SUA), dar și informal (Centre de dezbateri socio-politice, Universități Populare).

Au fost utilizate diverse forme de promovare a proiectului: versiunea demo a resursei lingvistice AnaDiP – primul soft realizat de studenta din anul III de la Facultatea de Informatică, Iulia Scutariu, prezentare video (film), prezentări electronice de tip .ppt, afișe, pliante, interviuri în presa ieșeană, clujeană, târgu-mureșeană, postarea unor lucrări pe diverse pagini online etc. Practic, nu a fost omisă nicio posibilitate de mediatizare a proiectului de cercetare, subliniindu-se de fiecare dată numele finanțatorului și numărul de contract. Acest program și-a găsit utilitatea nu numai în interpretarea contextului electoral, de la care s-a pornit inițial, dar și în înțelegerea și interpretarea (analiză cantitativă și calitativă) a contextului socio-politic într-un anumit moment (spre ex: perioada de pre-adoptare a Noului Cod al Muncii).

De asemenea, au fost citate rezultatele cercetării prezente în mai multe lucrări cu caracter științific, teza de disertație *Statistici lexicale ale discursului. Aplicație la analiza discursului public a Mădălinei Spătaru* susținută în iulie 2011, teza de doctorat *Analiza semiotică a limbajului politic eminescian*, susținută public în 27 septembrie 2011 și alte lucrări.

În acest fel, un asemenea instrument de analiză reușește să răspundă rapid și riguros la identificarea tonalităților discursive pe problematici politice și jurnalistice (partea retorică, dar și cea rațională, într-un discurs), necesare cercetătorilor din domeniul NLP sau specialiștilor din domeniul politic, sociologic, media.