



UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

## METODOLOGIA ANALIZEI AUTOMATE A DISCURSULUI (PROCESUL BOLOGNA)

### DESPRE CORPUS

Spre deosebire de unele abordări de *data mining* care au la baza extragerea informației dintr-o bază nestructurată de texte, studierea Procesului Bologna de către grupul nostru de lucru se face plecând, în primul rând, de la stabilirea corpusului documentar. Ne propunem astfel o structură de corpusuri pe mai multe paliere și care vor putea fi analizate fie individual, fie zonal (pe niveluri), fie global (reunite). Motivele unui asemenea demers țin de pertinența informației pe care dorim să o extragem în funcție de orientările interpretative prezentate în cele ce urmează.

- I. Un prim corpus numit de noi și corpusul central este reprezentat, aşadar, de documentele Procesului Bologna:
  - A) În centrul sferei documentare am situat, după cum rezultă din bibliografia consultată, **Declarația de la Bologna** (1999), **Conferințele ministeriale** (2001, Praga – 2012, București), precum și cele trei **forumuri** (2009-2012). Acestea trebuie studiate atât pentru degajarea orientărilor fundamentale, cât și pentru punerea în evidență a unei cronologii conceptuale care să poată reflecta istoricul acestui proces și ordinea de intrare a conceptelor în scenă.
  - B) Un al doilea corpus, reprezentat de **rapoartele naționale**, este organizat pe țări și poate permite analize comparative, dar și analize globale având la bază glosarul degajat din documentele fundamentale.
  - C) O altă categorie de documente o constituie **seminariile** (recente) analizate pe baza documentației de arhivă virtuală.
  - D) Vom lua în considerație documentele rezultate din întâlnirile **BFUG** (*Bologna Follow-up Group*). Vor analizate documentele grupurilor de lucru (WG) și care apar organizate tematic (*Social Dimension, Mobility, Lifelong learning, Employability, Qualification frameworks/ Degree structures, Education, research and innovation, International openness, Data Collection / Reporting on the Bologna Process Implementation, Financing and Governance of Higher Education, Quality assurance, Recognition, Transparency tools, Student Centred Learning*)<sup>1</sup>.
  - E) O serie de „sateliți” provin din documentele semnate de **membri consultativi** (Consiliul European, EUA, EURASHE, ESU, UNESCO, Education International, ENQA și BUSINESSEUROPE) referitoare la Procesul Bologna.
- Trebuie subliniat faptul că structurarea documentației din corpusul central s-a făcut în acord cu viziunea oficială de pe pagina EHEA (*European Higher Education Area*)<sup>2</sup>, entitate creată în 2010, ca emanație a acestui proces.
- II. În afara corpusului central prezentat mai sus, managementul proiectului a decis explorarea altor corpusuri care se referă cu precădere la acest proces, sau la situația învățământului

<sup>1</sup> Vezi *Work Programme* pe <http://www.ehea.info/>.

<sup>2</sup> Pe pagina BOLOGNA PROCESS – EUROPEAN HIGHER EDUCATION AREA (<http://www.ehea.info/>) găsim o structură de arhivă a documentației fundamentale: *Main documents (Ministerial conferences, Bologna Policy Forums), National reports, Recent seminars, BFUG Meeting documents*.





UNIUNEA EUROPEANĂ  
Fondul Social European



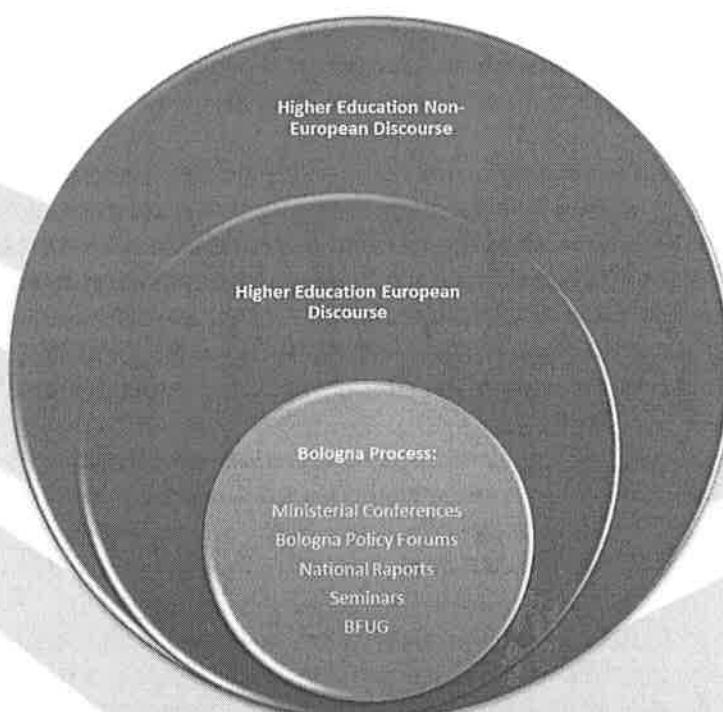
GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013 superior din EHEA (de exemplu, documente OCDE). Se încercă, în acest fel, dincolo de lărgirea bibliografiei, înțelegerea impactului pe care Procesul Bologna îl are asupra învățământului superior în prezent, dar mai ales în viitor, precum și înțelegerea cadrului european largit de dezbatere. Menționăm că EHEA este formată actualmente din 47 de state.

- III. Vor mai fi luate în considerație, documente elaborate de alte state neeuropene pentru a pune în evidență perspective contrastive și diferențiatoare care pot preocupa societatea globală. Atenția se îndreaptă asupra celor 19 state non-EHEA, ale căror abordări reprezintă un reper bibliografic și de studiu permanent.



Odată structurat corpusul, după un model arborescent, atât elaborarea dicționarului (care emerge din corpus), cât și analizele pot urma diferite linii de cercetare la nivel de microunivers sau de microunivers discursiv. Așa cum am arătat, acestea pot fi comparative și longitudinale.

O astfel de viziune este în acord cu perspectiva actuală din discursologie, în care abordările analitice trebuie să urmeze o evoluție ascendentă, de la text (discursul textual) la baze textuale tematice (discursul domeniului). Pentru o corectă interpretare, vom ține seama, pe de o parte, de apartenența textelor la același grup tematic omogen, precum și de data elaborării documentului. În anumite cazuri, se pot cerceta diversele grade de similaritate, adică de convergență tematică. Vom degaja, de asemenea, hărți ale cunoașterii și grafice obținute pe baza tratamentelor statistice.

Pentru crearea unei biblioteci esențiale pentru zonele a II-a și a III-a, referitoare la învățământul superior în EHEA și în statele avansate din afara Europei este necesară consultarea unor entități specializate, dar și a unor reviste de prestigiu internațional. Documentele vor fi descărcate într-un spațiu rezervat și ulterior filtrate, „catalogate” și folosite pentru viitoare consultări, analize etc.



UNITATEA EXECUTIVĂ PENTRU  
FINANȚAREA ÎNVĂȚĂMÂNTULUI  
SUPERIOR, A CERCETĂRII,  
DEZVOLTĂRII SI INOVĂRII



Politici publice fundamentate  
în Învățământul Superior



UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013 Este necesar ca, o dată grupate, să se procedeze la o atență examinare a tuturor textelor, indiferent de zona de proveniență. Este vorba de verificarea integrității fișierelor, a acurateței conversiei caracterelor de către programele informatiche de indexare semantică și de analiză care alcătuiesc **setul de instrumente informative (Tropes și Zoom/Owledge)**. Documentele sunt, în mare parte lor majoritate, în format PDF, ceea ce impune compararea conversiei atât cu IFILTER-ul produs de Adobe, dar și de Foxit, în multe cazuri mai performant.

Un alt concept metodologic este acela de „puritate” a textelor (eliminarea sau măcar diminuarea zgomotului informațional). Astfel, este de dorit ca o serie de metadate editoriale, abundente pe coperte și în primele pagini, bibliografia, indicii de autori și tematici să fie *provizoriu* eliminate atunci când dorim să obținem, prin analize, statistici conceptuale riguroase. În acceptia noastră, discursul nu poate fi corect pus în evidență decât dacă ne mărginim la analiza textului propriu-zis (eliminând datele bibliografice, editoriale, și care țin de tradiția tipografică). Multii analiști merg mai departe atunci fac analize lexicometrice privind ocurența absolută, preferând să substituie procedeele anaforice prin concepții la care se referă, pentru ca numărătoarea automată să nu fie viciată. Desigur, acest grad extrem de acuratețe, valabil pentru volumetrii rezonabile, nu este aplicabil în cazul nostru, întrucât volumul de informații este suficient de mare pentru a garanta fiabilitatea calculelor statistice.

### DESPRE ANALIZA AUTOMATA A DISCURSULUI BOLOGNA (asistată de *Tropes, Zoom/Owledge*)

Înțelegerea complexității analizei discursului bazată pe programul *Tropes*, principalul nostru instrument de lucru și care a fost adaptat acestui proiect, presupune cunoașterea întregului său univers operațional. Rezultatele unor analize pot părea, la prima vedere, dispensabile, dar înlănțuirea acestora, aşa cum putem vedea din schema de mai jos, formează o succesiune logică, necesară și armonioasă, concepută pentru a controla toate informațiile pe care le-am putea extrage dintr-un text, indiferent de dimensiunea acestuia.



UNIUNEA EUROPEANĂ  
Fondul Social European

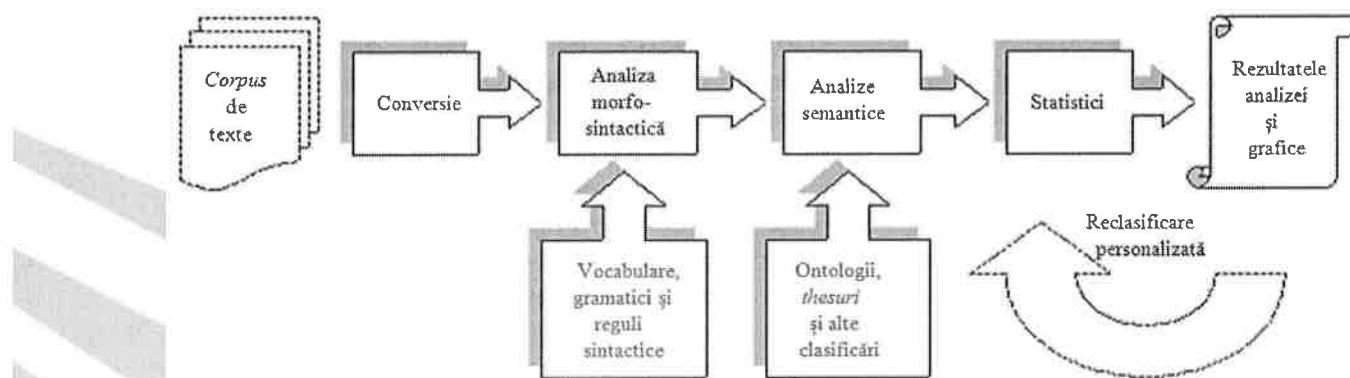


GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013



**Analiza morfosintactică.** După conversia textului, *Tropes* identifică categoria morfologică a tuturor cuvintelor din text, tratând în mod convenabil omonimele morfologice de tipul „may”, verb, și „may”, substantiv referitor la o lună a anului (se estimează la 2% rata medie de eroare).

Gramatica programului cuprinde următoarele clase morfologice:

- *Substantive* (fiind părți de vorbire purtătoare de informație pertinentă, acestea apar atât în meniul *Substantive*, cât și grupate în clase semantice de termeni echivalenți numite *Referințe*, după cum vom vedea mai detaliat, în cele ce urmează).
- *Verbe*, împărțite în *factive*, care exprimă acțiuni, *stative*, care exprimă stări sau posesia, și *reflexive*, care exprimă gânduri despre lume, sentimente.
- *Conectori* (conjuncții, locuțiuni conjuncționale, pronume relative), care introduc *condiția*, cauza, scopul, adunarea, disjuncția, opozitia, comparația, timpul și locul. Trebuie remarcat faptul că unele din aceste categorii reunesc și ale categorii de conectori și au la bază o clasificare proprie, justificată de autori<sup>3</sup>.
- *Modalizări*, clasificate ca exprimând: *timpul*, *locul*, *modul*, *afirmația*, *îndoiala*, *negația* și *intensitatea*. Aici sunt incluse în special adverbele și locuțiunile adverbiale.
- *Adjective*, împărțite în *subiective*, care exprimă proprietăți percepute subiectiv, cum ar fi „mare”, „important” etc., *obiective*, care exprimă proprietăți percepute obiectiv, cum ar fi: „instituțional”, „educativ” etc., și adjective *numerale*.
- *Pronume personale*.

O ultimă categorie, cea a *determinanților* (articole, prepoziții, unele pronume), nu are statut morfologic independent, dar o putem detecta prin plimbarea mouse-ului pe textul analizat.

<sup>3</sup> Ghiglione, R.; Landré, A.; Bromberg; M.; Molette, P.: *L'analyse automatique des contenus*, Paris, Dunod, 1998.





Proiect cofinanțat din Fondul Social European, prin Programul Operational "Dezvoltarea Capacității Administrative", în perioada 2007-2013 *Tropes* prezintă toate aceste clasificări, indicând numărul de ocorențe și procentajul în raport cu categoria principală. Am ales, spre exemplificare, un text aleatoriu (*Creativity in Higher Education*) unde verbele stative, de exemplu, reprezintă 28,1%, iar adverbele de timp 12,9%. În paranteze apare numărul de ocorențe, adică numărul de forme grupate în aceeași clasă, inventariate de program. În ferestrele din dreapta, marcate cromatic, pot fi observate câțiva din termenii astfel identificați.

<input checked="" type="checkbox"/> Stative 28.1% (617) <input checked="" type="checkbox"/> Reflexive 24.3% (533) <input checked="" type="checkbox"/> Performative 0.0% (0) <input type="checkbox"/> * Connectors : <input type="checkbox"/> Condition 2.7% (32) <input type="checkbox"/> Cause 4.4% (52) <input type="checkbox"/> Goal 2.1% (25) <input type="checkbox"/> Addition 60.9% (723) <input type="checkbox"/> Disjunction 6.6% (79) <input type="checkbox"/> Opposition 7.3% (87) <input type="checkbox"/> Comparison 13.3% (158) <input type="checkbox"/> Time 2.7% (32) <input type="checkbox"/> Place 0.0% (0) <input type="checkbox"/> * Modalities : <input type="checkbox"/> Time 12.9% (103) <input type="checkbox"/> Place 13.4% (107) <input type="checkbox"/> Manner 37.3% (297) <input type="checkbox"/> Assertion 2.8% (22) <input type="checkbox"/> Doubt 1.1% (9) <input type="checkbox"/> Negation 11.3% (90) <input type="checkbox"/> Intensity 21.2% (169) <input type="checkbox"/> * Adjectives : <input type="checkbox"/> Objective 69.2% (1565) <input type="checkbox"/> Subjective 13.5% (306) <input type="checkbox"/> Numerical 17.3% (391) <input type="checkbox"/> * Pronouns : <input type="checkbox"/> "I" 3.2% (6) <input type="checkbox"/> "Thou" 0.0% (0)	<ul style="list-style-type: none"> <li>EUA invited 32 higher education institutions from 21 countries representing diverse disciplinary foci and missions (art, technical)</li> <li>Main findings and conclusions: The project outcomes strongly support the notion that the participation of representatives from various countries in the network was key to the success of the project.</li> <li>This finding is noteworthy since values and ethical principles have received relatively little attention in the HEI reform and development processes.</li> <li>Values are of particular importance in a political climate in which many European universities are under considerable and increasing pressure to demonstrate their relevance and research are the very essence and major strength of higher education. Moreover, these principles reflect values which are important (in a material sense) of HEIS' activities.</li> </ul>
---	---

Analiza morfosintactică ne interesează în măsura în care dorm o corectă detectare a formelor compuse (substantive compuse, sintagme, nume de instituții, de programe etc.), precum și a acronimelor. Extractorul terminologic ne va ghida în acest sens, pentru ca întregul vocabular de interes să poată fi detectat și, ulterior, clasificat semantic.

**Analiza lexico-semantică.** O dată găsită unitatea semnificativă de bază (adică propoziția), *Tropes* va clasifica *referințele* (substantivele) în clase semantice, pe trei niveluri. Prin *referințe* înțelegem substantivele purtătoare de informație pertinentă. În aceasta etapă, cea mai complexă în tratamentul informației, programul se sprijină pe rețelele semantice proprii și pe dezvoltări speciale (clasificările noastre, numite *scenariu*). *Tropes* este pregătit să trateze ambiguitățile semantice cunoscute, calculând probabilitatea de ocorență a unui anumit sens într-un context particular. De exemplu, *Tropes* distinge «mouse», rozător, de «mouse», hardware. Acest tratament nu este întotdeauna ușor, pentru că, în realitatea textuală, este firesc ca diferite contexte opuse să se activeze în același timp. Rata medie de eroare nu depășește 5%.

References	References fields 2	References fields 1
------------	---------------------	---------------------





Proiect cofinanțat din Fondul Social European, prin Programul Operational "Dezvoltarea Capacității Administrative", în perioada 2007-2013

<input checked="" type="checkbox"/> 0179 creativity	<input type="checkbox"/> 0128 education	<input type="checkbox"/> 0343 education
<input type="checkbox"/> 0123 education	<input type="checkbox"/> 0125 telecommunication	<input type="checkbox"/> 0209 communication
<input type="checkbox"/> 0120 network	<input type="checkbox"/> 0114 scholar	<input type="checkbox"/> 0175 organization
<input type="checkbox"/> 0109 project	<input type="checkbox"/> 0105 person	<input type="checkbox"/> 0120 cognition
<input type="checkbox"/> 0101 collaborator	<input type="checkbox"/> 0099 educational_institution	<input type="checkbox"/> 0102 europe
<input checked="" type="checkbox"/> 0099 hei	<input type="checkbox"/> 0081 institution	<input type="checkbox"/> 0098 social_group
<input type="checkbox"/> 0094 scholar	<input type="checkbox"/> 0079 difference	<input type="checkbox"/> 0085 time
<input type="checkbox"/> 0081 institution	<input type="checkbox"/> 0066 time	<input type="checkbox"/> 0080 business
<input type="checkbox"/> 0080 university	<input type="checkbox"/> 0060 location	<input type="checkbox"/> 0073 location
<input type="checkbox"/> 0074 procedure	<input type="checkbox"/> 0059 knowledge	<input type="checkbox"/> 0054 characteristic
<input type="checkbox"/> 0058 knowledge	<input type="checkbox"/> 0055 organization	<input type="checkbox"/> 0044 foreign_text
<input checked="" type="checkbox"/> 0055 higher_education_institution	<input type="checkbox"/> 0042 french_text	<input type="checkbox"/> 0043 behavior
<input checked="" type="checkbox"/> 0044 euia	<input type="checkbox"/> 0041 creation	<input type="checkbox"/> 0042 politics
<input type="checkbox"/> 0042 french_text	<input type="checkbox"/> 0041 partnership	<input type="checkbox"/> 0041 creation
<input type="checkbox"/> 0041 investigation	<input type="checkbox"/> 0040 cognition	<input type="checkbox"/> 0033 security
<input type="checkbox"/> 0041 partnership	<input type="checkbox"/> 0036 government	<input type="checkbox"/> 0032 society
<input checked="" type="checkbox"/> 0039 activity	<input type="checkbox"/> 0033 exploitation	<input type="checkbox"/> 0031 language
<input type="checkbox"/> 0037 difference	<input type="checkbox"/> 0033 staff	<input type="checkbox"/> 0026 city
<input type="checkbox"/> 0036 dissimilarity	<input type="checkbox"/> 0032 europe	<input type="checkbox"/> 0023 device
<input type="checkbox"/> 0035 government	<input type="checkbox"/> 0032 society	<input type="checkbox"/> 0020 mechanics
<input type="checkbox"/> 0034 report	<input type="checkbox"/> 0032 west_europe	<input type="checkbox"/> 0020 art
<input type="checkbox"/> 0033 development	<input type="checkbox"/> 0030 mechanism	<input type="checkbox"/> 0019 media
<input type="checkbox"/> 0033 staff	<input type="checkbox"/> 0029 important_person	<input type="checkbox"/> 0019 agriculture
<input type="checkbox"/> 0032 society	<input type="checkbox"/> 0029 way	<input type="checkbox"/> 0017 feeling
<input type="checkbox"/> 0028 context	<input type="checkbox"/> 0026 city	<input type="checkbox"/> 0017 work
<input type="checkbox"/> 0028 structure	<input type="checkbox"/> 0026 sameness	<input type="checkbox"/> 0014 money
<input type="checkbox"/> 0028 potential	<input type="checkbox"/> 0024 social_group	<input type="checkbox"/> 0014 entertainment
<input type="checkbox"/> 0027 practice	<input type="checkbox"/> 0021 security	<input type="checkbox"/> 0013 science

Așa cum putem vedea în exemplul de mai sus, în analiza semantică referințele se grupează ulterior în două câmpuri, ceea ce ne permite înțelegerea temelor abordate. Aceasta funcție este deosebit de importantă în indexarea documentară (detectarea automată a descriptorilor, a temelor). De exemplu, dacă un text menționează „United Kingdom, Britan și England”, *Tropes* deduce că în text se vorbește despre „United Kingdom”, care este o țară din „Europa”. În esență, este vorba de contragerea semantică aplicată de program. Sunt calculate astfel numărul de ocurențe, iar vizualizarea se poate face atât alfabetic, cât și în ordinea descrescătoare a ocurențelor (ca în exemplul de mai sus). Cuvintele tematice cele mai des menționate vor ocupa primele locuri.

Pe lângă aceasta clasificare, *Tropes* pune în evidență, în mod automat, referințele care apar deseori *relaționate*. În exemplul nostru, „creativity” și „higher education” apar cel mai frecvent relaționate, ceea ce indică insistența autorului asupra acestor două concepte pe care le regăsim și în titlu.

0022 [creativity > higher education]	European commission, and staff development and reward schemes were identified as key factors for fostering the human potential for creativity in higher education. diverse learning experiences will likely result in conditions favourable to the creativity of the higher education community. national and European levels may exert considerable influence on creativity within the higher education sector. Governments need to be aware of the responsibilities this entails. Yet successfully fostering creativity in higher education will most probably provide Europe with a competitive advantage in its endeavour to develop a knowledge based economy in the world. Creativity in Higher education, a project initiated by the European university association (EUA) and co-financed by the Societas Programme of exploratory activity to enhance our understanding of the concept. and to address the question of how creativity can be strengthened in European higher education. In this project the term creativity has been originally coupled with appropriateness in all disciplines including the arts. usually in higher education may be enhanced... (or hindered) by specific institutional and environmental situations as well as cultural factors. Internal and external factors influencing creativity in HEIS 2. CONCEPTUAL FRAMEWORK Economic/social environment, infrastructure (p. 9)
0025 [projects > partner]	
0016 [networks > reports]	
0016 [creativity > agendas]	
0016 [creativity > projects]	
0012 [knowledge > creation]	
0011 [higher education > networks]	
0010 [higher education > sector]	
0009 [literature > creativity]	
0009 [quality assurance > agencies]	
0009 [dimension > creativity]	
0008 [teaching > learning]	
0008 [higher education > partner]	
0008 [partner > institutions]	
0007 [innovation > nature]	

*Tropes* analizează relațiile fiecărei referințe utilizate arătând modul cum rețelele apar întreținute în fiecare text. În acest caz textul vorbește despre „creativitate”. *Tropes* arată, rapid și fiabil, relațiile conceptului





UNIUNEA EUROPEANĂ  
Fondul Social European

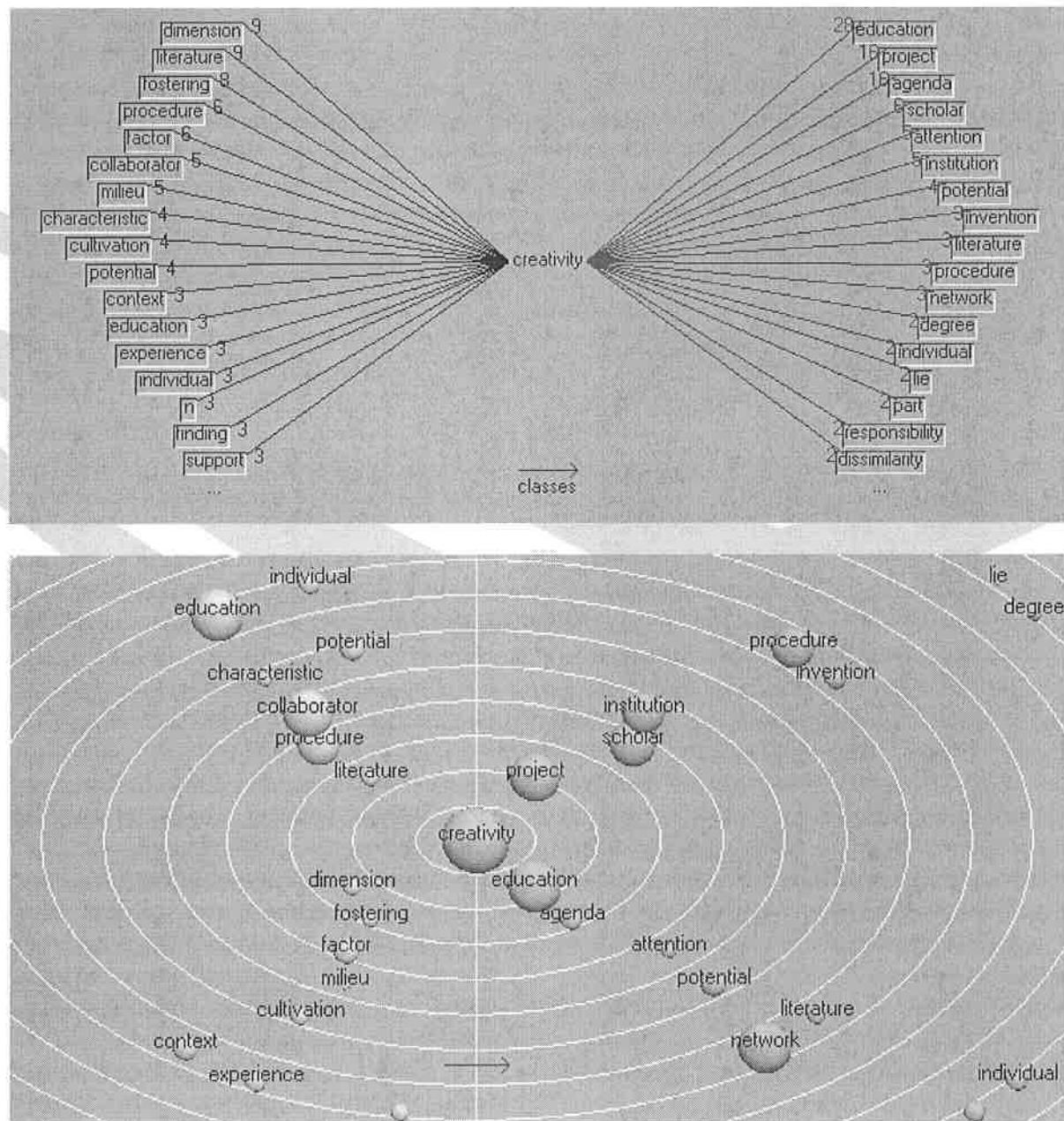


GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013 central cu toate referințele despre care autorul pomenește. Un grafic în stea este afișat pentru a înclesni percepția legăturilor. Aceeași informație poate fi vizualizată și într-un grafic de sfere (pe modelul sistemului solar). În mod interactiv, utilizatorul poate cerceta și alte configurații relationale, selecționând oricare din aceste obiecte (concepte).



Să ne imaginam un analist care își pune următoarea întrebare: „Cum anume relaționează autorul «creativitatea» cu «literatura»?” Tropes răspunde afișând o frază precum aceasta:

*uefiscadi*

UNITATEA EXECUTIVĂ PENTRU  
FINANȚAREA ÎNVĂȚĂMÂNTULUI  
SUPERIOR, A CERCETĂRII,  
DEZVOLTĂRII SI INOVĂRII



Politici publice fundamentate  
în Învățământul Superior



UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice

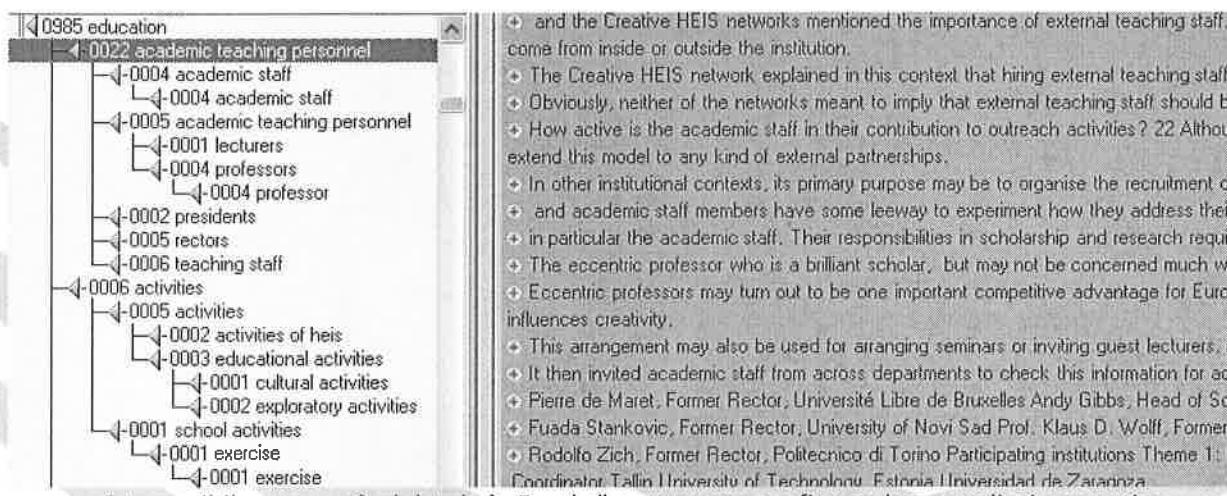


Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Project cofinanțat din Fondul Social European, prin Programul Operational "Dezvoltarea Capacității Administrative", în perioada 2007-2013

- Similarly to many other aspects of creativity, there is a great deal of literature on the development of creativity in the context of human resources
- In contrast, the literature also suggests that monitoring, evaluation and support when geared towards improvement can enhance the intrinsic motivation in individuals that is considered highly favourable to creativity.

Trebuie să subliniem că utilizatorul programului are deplina libertate de a folosi propriile sale clasificări, aşa cum am procedat și noi, construind un *scenariu* propriu și modificându-l constant în funcție de obiectivele cercetării. În exemplul următor, vom folosi un *thesaurus* dezvoltat în acest proiect (a se vedea raportul referitor la setul de instrumente) pentru a avea acces la arborescența conceptuală.



Este posibil ca exportând datele în Excel să putem crea grafice pe baza analizei semantice:





UNIUNEA EUROPEANĂ  
Fondul Social European

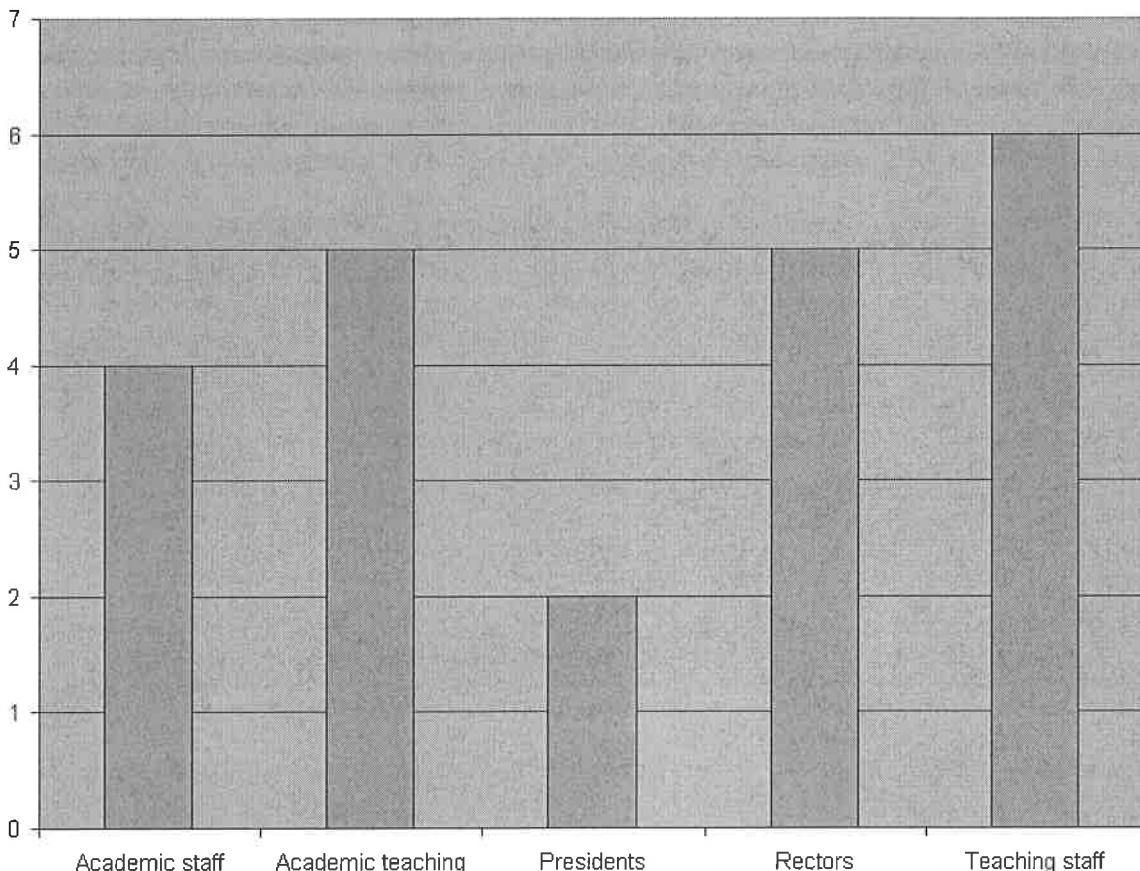


GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013  
Academic teaching personnel



Astfel, cele patru analize prevăzute în cadrul proiectului vor pune în evidență datele de natură conceptuală tratate din perspectivă statistică. Tratamentul va fi realizat atât cu Excel, cât și cu programe specifice, cum este Gephi.

Al doilea program, legat de *Tropes*, numit *Zoom* ne permite indexarea semantică a unui corpus textual. Așa cum am menționat atunci când ne-am referit la corpusul selecționat pentru analize, baza textuală pe care o avem în vedere este vastă. Prin urmare, după dezvoltarea instrumentelor lingvistice necesare proiectului cu ajutorul programului *Tropes*, când dicționarele sunt definitiv reglate în funcție de obiectivele analizelor pe care dorim să le elaborăm, vom recurge la indexarea semantică cu *Zoom*, pentru a înțelege, de astă dată, care texte, considerate global, sunt pertinente din perspectiva analizelor dorite. De asemenea, în această etapă se pot ridica diverse chestiuni legate de similaritatea textuală, de natură mai ales comparativă.



Proiect cofinanat din Fondul Social European, prin Programul Operational "Dezvoltarea Capacitatii Administrative", in perioada 2007-2013

The figure shows a Windows desktop environment. In the foreground, a 'File Types' dialog box is open, listing various file extensions and their associated programs. The 'Search' tab is selected, and the results for '\*.pdf' are displayed. The results list several PDF documents, including titles like '00000 VA Changed Economy with Unchanged Universities.pdf', '00000 VA Sustainable Future for the European Higher Education Area 2010 to 2020 - The View of Higher Education Researchers.pdf', and '00000 VA Test of Leadership - Charting the Future of U.S. Higher Education.pdf'. Below the search results, a note states: 'The Scenario shows a classification of the files found. Use Tropes to create or customize your classifications.'

Programul prezintă marile teme abordate de diversele corpusuri. Numerele din paranteze indică câte texte din acest corpus abordează o anumită temă. Putem observa că 120 de texte vorbesc despre „sănătate”, dintr-un total de 141. Această funcție permite filtrarea textelor care ne interesează cu adevărat, putând merge până la detaliu, prin înmulțirea criteriilor.

În exemplul următor, putem observa că numai 53 texte vorbesc de „creativitate” din totalul textelor luate în considerație.

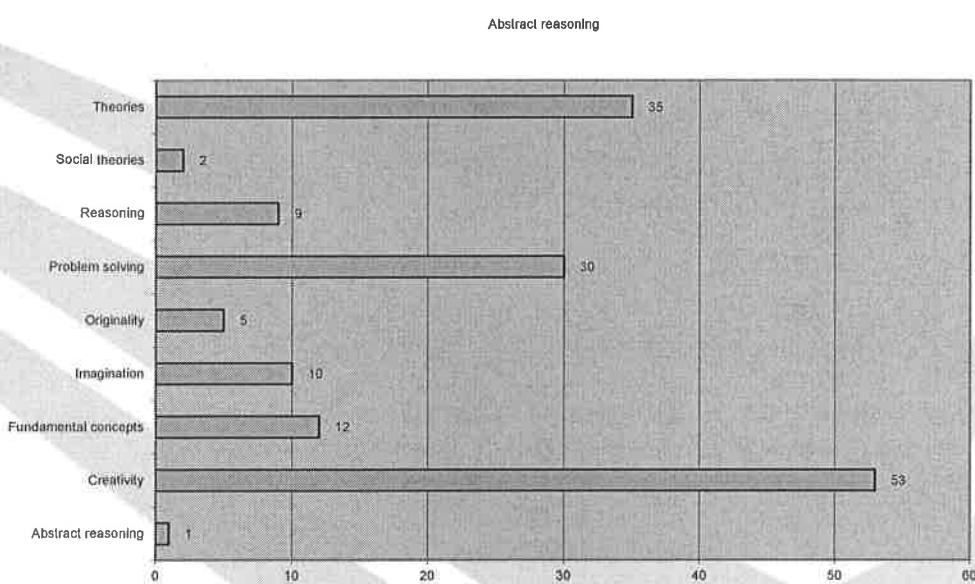
The screenshot shows the WinRAR search interface. The search term 'creativity' has been entered in the search bar at the top. Below the search bar, there is a 'Folder' button and an 'Index this folder' checkbox. The main area displays a list of search results under the heading 'Files found'. The results are as follows:

- 00179 Creativity in Higher Education.pdf [493 kb] - A report from the EUA Creativity Project, dated 2006-2007, containing information on creativity in higher education.
- 00012 EUA - Annual Report 2007.pdf [2763 kb] - The annual report for 2007 from the European University Association.
- 00005 The Bologna Club - What U.S. Higher Education Can Learn from a Decade of European Reconstruction.pdf [740 kb] - A report from the Bologna Club discussing higher education reform in the United States based on European experiences.
- 00005 Territory Education for the Knowledge Society. Volume 1.pdf [5095 kb] - A large document related to territorial education for knowledge societies.
- 00004 Policy Making, Strategic Planning, and Management of Higher Education.pdf [333 kb] - Papers on higher education policy making, strategic planning, and management.
- 00002 Work-based learning. Illuminating the Higher Education Landscape.pdf [593 kb] - A report on work-based learning in higher education.



Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013 Așa cum se poate remarcă, textul cel mai important este cel ales de noi. Acesta ocupă primul loc, cu 179 de ocorențe. Pe locul al doilea, se află un alt text, cu numai 12 ocorențe. Astfel, cercetătorul nu va continua analiza, restul fiind dispensabil. Aceasta funcție de filtrare este prețioasă mai ales când numărul de texte de analizat este foarte mare și când dorim să scoatem în evidență textele cu adevărat semnificative („top 10”, să spunem)

În Zoom putem căuta folosind un simplu concept sau o întreagă clasă semantică. Programul dispune și de alte funcții importante pentru analist, cum ar fi căutarea unor texte similare cu un anumit text, funcționalitate care ne interesează în mod particular. Desigur, putem produce statistici plecând de la rezultatele obținute.



Zoom poate fi substituit de un alt program (aflat încă în prototip), Owledge, care ne poate duce mai departe spre clusterizarea corpusului, pentru a observa aglomerări conceptuale, relații frecvente în întregul univers textual.

Owledge nu oferă date interesante din punct de vedere statistic dacă corpusul este prea mic sau dacă este excesiv de omogen. Altfel spus, prin Owledge putem surprinde ceea ce nu este banal, repetitiv, în configurațiile conceptuale. În bună parte, acest software servește „descoperirii”, adică ne ajută să revelăm legături conceptuale la care poate nu ne-am gândit.

În exemplul de mai jos, am folosit un corpus suficient de mare. Este vorba de o biblioteca despre învățământul superior, din care face parte și textul analizat, aflată în faza de filtrare pentru zonele a II-a și a III-a ale cercetării noastre. Am ales concepțele „creativity” și „higher education” pentru o clusterizarea informației în jurul acestei idei. De astă dată, nu ne interesează concepțele din text, ci informația din corpus. Rezultă un grafic interesant, din care prezentăm mai jos o bună parte:



Proiect cofinanțat din Fondul Social European, prin Programul Operational "Dezvoltarea Capacității Administrative", în perioada 2007-2013

living standards	innovativeness	illiteracy	kindergarten	study grant	educational effect	central asia
structural adjustment	bachelor of arts	inclusive education	rote learning	performance-based	academic evalua	east africa
intellectual work		demand for teachers	interactive learning	study loans	admission to exam	hunger
financial allocation	technical colleges	training needs	prior learning	secondary school le	oral examination	cognitive ability
	african universities	educational sciences	academic education	secondary school st	recognition of prior	aggression
social capital	educational psycho	philosophy of educatio	levels of training	special schools		work environment
adjunct professors	elective courses	mediator	postcompulsory educati	acceleration	stimuli	
teacher educators	optional courses		postgraduate study	numerus clausus	initial teacher educ	personal growth
computer games	refresher course	educational models		gifted students	secondary school	individual developm
extracurricular activities	university curriculum	curriculum planning	university planning	accreditation of pro	animation	social relations
administrative bodies			academic management	accreditation proced	teaching units	individualism
academic authority	punishment	formative evaluation	project implementation	accreditation proced	learner centred ap	skill development
	lifelong education	quota system	violence	accreditation status	standardised tests	visualization
	community education		natural disaster	external accreditatio	conflict resolution	
curriculum reform	career education	academic autonomy	professional profile	assessment proced	technology educa	bilateral agreement
	education courses		habilitation			law enforcement
		curriculum framework		attestation	management trainin	citizen participation
decentralization	educational infrastruc	educational plans	corruption	code of professiona	mongolia	city government
	quality of the higher	educational practice		professional compet	central america	provincial governme
educational innovations	educational theories					

Desigur, cercetătorul va asocia și alte concepte care îl interesează, dacă dorește aprofundarea pe mai departe a analizei. Aceasta îl va conduce la o nouă filtrare a textelor, până se vor depista acele texte care să satisfacă toate criteriile de căutare.

Vom încheia aici prezentarea tipurilor de analiză în funcție de programele care ne stau la dispoziție, pentru a ne ocupa ulterior de chestiunile metodologice propriu-zise.





UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

## REPERE METODOLOGICE PRIVIND ANALIZA AUTOMATĂ A DISCURSULUI ȘI INTERPRETAREA ACESTEIA

Dacă referitor la corpus socotim că au fost precizate condițiile care ne permit selectarea, structurarea și tratamentul necesar atât elaborării dicționarului de lucru (a se vedea raportul referitor la setul de instrumente), cât și analizelor, în privința acestora din urmă se mai impun o serie de specificări de natură metodologică.

Așa cum am arătat în secțiunea anterioară, prin analizele automate încercăm, în primul rând, să stabilim frecvența termenilor (referințelor) în diversele corpusuri pentru a putea evidenția conceptele cheie. Acest lucru este posibil în măsura în care dicționarele reflectă în mod corect realitatea textuală, altfel spus, înregistrează toate cuvintele și sintagmele care formează fluxul discursului.

După degajarea celor mai importante concepte, fie la nivelul unui singur corpus, fie la nivelul marelui corpus, este firesc să încercăm să precizăm relațiile dintre aceste concepte. Relațiile de care vorbim sunt abstracte, scoase din context, și sunt puse în evidență de program la nivel de macrounivers. Vor fi, în felul acesta, observate relațiile cele mai frecvente, dar și configurația spațială, precum și tendințele de aglomerare.

Putem afirma, că o dată stabilite conceptele și relațiile dintre acestea, calculate ocurența și frecvența legăturilor, am putea pune în lumină întregul schelet discursiv. Firește, atunci când vorbim de ocurențe și de frecvență relațiilor avem în vedere, de la analiză la analiză, un anume prag de detectare a relațiilor pentru a evita nespecificitatea. Întregul demers este, aşadar, un drum spre esența discursului. Vrem să stim astfel ce se spune și, mai ales, cum se spune atunci când se vorbește de învățământul superior, bunăoară.

Este bine de subliniat faptul că anumite corpusuri sunt exhaustive, ca în cazul conferințelor ministeriale (8) sau al forumurilor de politică Bologna (3). Aceasta însemnă că centrul sferei de documente este inchetionabil și poate fi tratat adecvat de analist. La fel de lipsă de stăm și în cazul rapoartelor naționale. În 2001, 33 de state formau ceea ce ulterior va numi spațiul Bologna. În 2003, la Berlin, se ajunge la 40 de state membru, pentru ca la Londra, în 2007, să avem 46 de state, iar la Budapesta și Viena, 47. Aceasta este amplitudinea EHEA, desigur peste limitele geografice europene, ultimul stat admis fiind Kazahstanul. Principala sarcină a analizei este, din punctul nostru de vedere, ordinea intrării conceptelor în scenă și protagonismul actorilor. În concret, dorim să stim când s-a vorbit pentru prima dată oficial de „învățarea de-a lungul întregii vieți”, de exemplu, și ce state au aplicat primele programe în această direcție.

Neîndoios, zonele a II-a și a III-a nu pot fi, din perspectivă textuală, la fel de complet determinate. Este cu neputință să cercetăm tot ce s-a scris în Europa sau în afara ei asupra procesului Bologna. Este cu neputință să producem analize „intelibile” când volumul de date excede limitele umane. Este important de reținut însă că, în aceste cazuri, fiabilitatea este judecată în limita conținutului corpusului declarat și acest lucru trebuie precizat clar la fiecare analiză.

O altă direcție a analizelor o reprezintă dimensiune tematică, abordată de BFUG. Considerăm că acest tip de analiză va scoate în evidență o serie de concepte pe care se intemeiază o anumită tematică. De exemplu, ne putem întreba, atunci când vorbim despre dimensiunea socială, la ce se face în mod expres referință? Ce concepte presupune „construirea” acestei dimensiuni? De aceea, de astă dată, avem mai curând de-a face cu hărți conceptuale (*knowledge maps*) pentru o mai lipsă și mai sugestivă reprezentare a gândirii.



UNITATEA EXECUTIVĂ PENTRU  
FINANȚAREA ÎNVĂȚĂMÂNTULUI  
SUPERIOR, A CERCETĂRII,  
DEZVOLTĂRII SI INOVĂRII



Politici publice fundamentate  
în învățământul Superior



UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

Din unghiul de vedere al resurselor discursive (text și imagine) vom include, în analize, nu doar interpretarea analistului, dar și extrase semnificative din textele cele mai semnificative pentru o justă echilibrare a discursului textual cu cel imagistic.

În sfârșit, cercetătorii vor avea la dispoziție atât corpusurile, cât și instrumentele folosite în cazul în care altcineva ar dori reproducerea sau continuarea acestor analize în viitor. În ceea ce ne privește, ne vom limita doar la cele propuse în cadrul proiectului, revenind metodologic ori de câte ori pot surveni dificultăți particulare. Rapoartele acestor analize și rezolvarea problematicii întâlnite completează, în parte, prezenta încercare de metodologie generală.



UNITATEA EXECUTIVĂ PENTRU  
FINANȚAREA ÎNVĂȚĂMÂNTULUI  
SUPERIOR, A CERCETĂRII,  
DEZVOLTĂRII SI INOVĂRII



Politici publice fundamentate  
în Învățământul Superior



UNIUNEA EUROPEANĂ  
Fondul Social European



GUVERNUL ROMÂNIEI  
Ministerul Dezvoltării Regionale  
și Administrației Publice



Inovație în administrație  
Programul Operațional  
"Dezvoltarea Capacității  
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

## BIBLIOGRAFIE

- CARAGEA, Dan, „Analiza automată a discursului”, in Dumitrache, Ioan și Iovu, Horia (coord.), *Manual de autorat științific*, Editura Politehnica Press, București, 2011.
- GARRIC, Nathalie ; LEGLISE, Isabelle, „La place du logiciel, du corpus, de l'analyste : l'exemple d'une analyse de discours patronal à deux voix”, in G. WILLIAMS, coord., *Linguistique de corpus*, Presses Universitaires de Rennes, Rennes, 2005, pp. 101-113.
- GHIGLIONE, Rodolphe; LANDRÉ, Agnès; BROMBERG, Marcel; MOLETTE, Pierre, *L'analyse automatique des contenus*, Dunod, Paris, 1998.
- LANGLIE, Nicholas, „Educational Technology Leaders: Competencies for a Conceptual Age”, ISTE, 30th annual NECC, 2009.
- MARCHAND, Pascal ; MOLETTE, Pierre, „Convergences et spécificités de la statistique lexicale et de l'indexation syntaxico-sémantique automatique”, communication aux 4ème JADT, Nice, 1998.
- MARCHAND, Pascal, „Automatique (Analyse)”, in Patrick CHARAUDEAU ; Dominique MAINGUENEAU (dir.), *Dictionnaire d'analyse du discours*, Seuil, Paris, 2002.
- MARCHAND, Pascal, *L'Analyse du Discours Assistée par Ordinateur*, Armand Colin, Paris, 1998.
- MOLETTE, Pierre, „De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste”, conférence au colloque international "Psychologie Sociale et Communication", Tarbes, 2009.
- VAN DIJK, Teun A., *Discourse and Context. A sociocognitive approach*, University Press, Cambridge, 2008.

