

VERSIUNEA ON-LINE A DICȚIONARULUI ASOCIATIV ROMÂN

Victoria Bobicev, dr.conf.univ., Victoria Maxim, dr.conf.univ., Victoria Lazu, drd
Universitatea Tehnică a Moldovei

INTRODUCERE

Direcția de cercetare dedicată dezvoltării Inteligenței Artificiale (IA) rămâne una prioritară pe parcursul a multor ani. Pe când scopul final a ramurii date uneori pare să fie de neatins, dezvoltarea domeniilor adiacente ne permite să ne apropiem de rezultatul dorit din diferite direcții. Una din direcțiile de cercetare din domeniul inteligenței artificiale pe tot parcursul dezvoltării sale este Lingvistica Computațională care are ca scop modelarea limbajelor naturale și dezvoltarea sistemelor capabile să le prelucreze în modul similar cu cel al oamenilor. Direcția aceasta de cercetare folosește teoriile și descoperirile din astfel de domenii ca filologia, lingvistica generală, pedagogie și psihologie pentru a înțelege modul în care oamenii învață și utilizează limba naturală. Însă, mediul de bază în care se dezvoltă lingvistica computațională este informatica fiind acest instrument, cu ajutorul căreia formalizăm modelele propuse de alte domenii și obținem produsul final în formă de aplicații. Mai mult ca atât, anume domeniul informaticii este acum acela care furnizează resursele necesare pentru dezvoltarea lingvisticii computaționale și Procesării Limbajului Natural (PLN) ca subramură a ei. Studiul eficient al limbii este imposibil fără resurse lingvistice și lexicale, resursele lexicale fiind dicționare, vocabulare, lexicoane, tezaure. Astfel de resurse sunt costisitoare și este nevoie de mult timp și muncă minuțioasă pentru a completa un dicționar. Însă dicționarele create pentru oameni sunt prea complicate pentru aplicațiile soft care prelucrează textul în limbaj natural. Astfel, devine esențial de a crea dicționare specifice, adaptate la necesitățile specifice domeniului procesării limbajului natural. Un dicționar organizat și codificat în forma accesibilă pentru aplicațiile soft este denumit lexicon computațional. Astfel de lexicoane sunt în format electronic, sunt codificate cu ajutorul instrumentelor speciale și conțin informația necesară aplicațiilor soft. O mare parte din informația stocată într-un lexicon computațional este de caracter semantic.

Semantica s-a dovedit a fi elementul de bază al limbajului. Rețelele semantice lexicale sunt de o mare importanță în lingvistica computațională din

zilele noastre. Popularitatea largă a WordNet-ului [1] este argumentul ce demonstrează utilitatea lexicoanelor semantice. Unul din neajunsurile WordNet-ului este numărul mic de relații semantice introduse între cuvintele lexiconului. Alte lexicoane semantice, cum ar fi EuroWordNet [2] și Simple [3] au fost create pentru a rezolva această problemă. Relațiile semantice din aceste lexicoane sunt bine considerate de către lingviști competenți și se bazează pe diferite teorii lexicale.

Crearea generației moderne a dicționarelor și lexicoanelor este ușurată de instrumente oferite de mediul informatizat. În primul rând, în ziua de azi volumul de texte în format electronic a devenit imens și marea parte din textele acestea poate fi accesată cu ușurință pentru a analiza orice cuvânt și utilizarea lui în practică. Acest lucru accelerează studiul lexicului limbii, descrierea și organizarea lui. În a doilea rând, fenomenul așa numitului Web 2 a demonstrat abilitatea utilizatorilor de a crea resurse noi valoroase. Exemple de astfel de resurse sunt Wikipedia¹ și Dexonline² care au fost create exclusiv de utilizatorii internetului. ConceptNet [4] și BabelNet [5] sunt exemple de lexicoane semantice create utilizând această metodă modernă de a obține informația necesară.

Modul în care privim dicționarele (crearea și utilizarea acestora) s-a schimbat în mod drastic în ultimii ani. În timp ce au fost considerate o anexă la gramatică în trecut, în prezent dicționarele au trecut pe planul central. Într-adevăr, cu greu găsim o sarcină în procesarea limbajului natural (PLN) care poate fi efectuată fără referință la ele. Dicționarele actualmente sunt privite nu numai ca entități statice (aspect de baze de date), dar și ca rețele dinamice, adică grafuri, ale căror noduri și legături (puncte de conexiune) se pot modifica în timp. Interesant este faptul, că proprietățile legate de tipologie, clasificare și evoluție, cunoscute din alte discipline (societate, economie, studiul creierului uman), se aplică, de asemenea, pentru dicționare: totul este legat, prin urmare accesibil, și totul este în evoluție. Astfel, crearea lexiconului prin interogarea directă a utilizatorilor ne permite să obținem o mai bună înțelegere în ceea ce privește lexicul mental și să

¹ <https://www.wikipedia.org>

² <https://dexonline.ro/>

integrăm aceste observări în aplicațiile ce prelucrează textele.

Am creat lexiconul nostru, bazându-ne pe câteva principii de bază. Sursa de relații este primul principiu de bază în lucrarea noastră. Relațiile dintre cuvinte sunt obținute direct de la vorbitorii nativi ai limbii prin asociațiile lor libere. Cel de-al doilea principiu este tipul relațiilor implicate. Noi n-am denumit aceste relații nici nu le-am clasificat; acestea sunt doar relații de asocieri libere în mintea umană. În psihologie asociațiile libere sunt primele cuvinte care apar în mintea unui vorbitor nativ atunci când îi este prezentat un cuvânt stimulent, recuperate probabil din memoria asociativă [6]. Cuvântul prezentat respondentului este numit „stimulent” și cuvântul care îi vine în minte este numit „răspuns”. Acest tip de relații este studiat în diferite domenii de cercetare, cum ar fi psihologia, inteligența artificială, lingvistica computațională și procesarea limbajului natural.

Dicționarul asociativ este o colecție de perechi de cuvinte „stimulent – răspuns” și reprezintă limbajul într-o formă oarecum neobișnuită - nu sub formă de text continuu, ca într-un roman sau un articol de ziar, nu sub formă de o descriere sistematică, ca într-o gramatică sau dicționar, ci ca perechi (combinări) de cuvinte sau grupuri de cuvinte care servesc drept material de construcție pentru formularea gândurilor în frazele și propozițiile corespunzătoare.

Lucrarea raportează despre un experiment asupra creării bazei de date cu asociații de cuvinte pentru limba română. În prima fază a experimentului asociațiile au fost colectate cu ajutorul anchetei - chestionar. Apoi a urmat a doua fază care include crearea interfeței online și augmentarea dicționarului prin internet. Sunt discutate o serie de aspecte tehnice ale fazei a doua. Sunt prezentate unele statistici preliminare și se face o analiză succintă a bazei de date obținute. La această etapă a lucrului suntem interesați de îmbogățirea bazei de date cu cele mai reprezentative asocieri de cuvinte; analiza detaliată este amânată pentru cercetările viitoare.

1. LUCRĂRI CONEXE

Există un număr de lexicoane semantice cu diferite relații dintre cuvinte. Cel mai popular este WordNet care conține un număr relativ mic de relații; acest lucru este considerat unul dintre dezavantajele sale. Autorii EuroWordNet au revăzut și au lărgit acest set de relații. Simple folosește teoria structurii Qualia ca o sursă de relații

semantice în lexicon [7]. Încercarea, însă, de a codifica cât mai multe relații posibile are și un efect negativ; aceste lexicoane sunt dificil de prelucrat. Sunt necesari algoritmi destul de sofisticati pentru a obține informația utilă într-un timp plauzibil.

O altă sursă de informații semantice sunt Bazele de cunoștințe. Binecunoscuta CYC [8] include lexiconul ca parte a bazei de cunoștințe. Cuvintele din lexicon sunt legate de conceptele din baza de cunoștințe, obținându-se astfel o capacitate semantică. Numărul de concepte și relații este unul dintre cele mai mari între diferitele resurse de acest gen. Dimpotrivă, ConceptNet descrie doar 20 tipuri de relații; unele dintre ele sunt similare cu alte resurse. Este una din primele resurse care nu este creată de lingviști calificați, ci de voluntari prin intermediul interfeței online. Această metodă de dobândire de cunoștințe are mai multe avantaje: nu este nevoie de lingviști profesioniști cu pregătire specială, astfel, aplicarea metodei date duce la costuri mai mici și rată de creștere mai mare pentru o bază de date lingvistice.

Asociațiile între cuvinte sunt obținute de la oamenii fără cunoștințe speciale de lingvistică; singura condiție este ca aceștia să fie vorbitori nativi de limbă. Deși experimentele de asociere de cuvinte sunt o practică psihologică obișnuită, rezultatele obținute sunt de mare interes în diverse domenii ale cercetării, ca, de exemplu în domeniul științei cognitive. Cel mai important între toate acestea este înțelegerea faptului că asocierea este un mecanism fundamental al cunoașterii umane [9, 10]. În domeniul procesării limbajului natural această noțiune este reprezentată prin aplicarea metodelor statistice la un corpus de texte. Astfel de experimente au fost numite modelarea statistică a textelor [11]. Una din metodele statistice care modelează legăturile între cuvinte este noțiunea de informație reciprocă ca o măsură a importanței unei asocieri între două cuvinte [12]. Rețelele lexicale, reprezentate de noduri lexicale [13] sunt punctele de bază ale multor modele de conexiune ale gândurilor omenești.

Recent, asocierile de cuvinte au fost studiate de o serie de cercetători în domeniul științelor cognitive [14]. Toate aceste studii folosesc asocierile de cuvinte, rimele și normele fragmentelor de text [6] ale Universității din Florida de Sud, care este cea mai mare bază de date a asocierilor de cuvinte pentru limba engleza americană, care cuprinde aproape 5000 de cuvinte și un număr mediu de asocieri de 149 pentru fiecare cuvânt colectate de la mai mult de 6000 de participanți pe parcursul anilor 1975-2000.

Există diverse surse de asocieri de cuvinte pentru diferite limbi. Cea mai mare bază de date de asocieri de cuvinte este cea deja menționată pentru limba engleză³. Trebuie să menționăm, de asemenea, Tezaurul asociativ din Edinburgh [15] - bază de date disponibilă gratis iarăși pentru limba engleză.⁴ Printre resursele pentru alte limbi merită să fie menționate: dicționarul asociativ rus [16], dicționarul asociativ din Bulgaria [17], dicționarul slavon integrat [18]. Toate aceste surse au fost colectate manual, utilizând chestionare. Cele mai recente surse au fost create folosind interfața online. Printre acestea sunt asocierile de cuvinte pentru cuvintele japoneze [19], dicționarul asociativ francez⁵, jocul de cuvinte - asocieri pentru limba engleză,⁶ interfața online pentru dicționarul asociativ rus.⁷

2. PRIMA ETAPĂ A CREĂRII BAZEI DE DATE A ASOCIERILOR DE CUVINTE ROMÂNEȘTI

Prima colecție de asocieri de cuvinte în limba română a fost creată prin interogare directă. 150 de cuvinte stimulente au fost selectate din lista celor mai frecvente cuvinte în limba română. Lista cuvintelor frecvente a fost creată pentru corpusul descris la [20]. Corpusul a fost creat pe baza a 93 de cărți de diferite genuri: ficțiune românească și străină, literatură religioasă, filosofie, texte medicale, istorie, drept și altele. Scopul autorilor a fost de a include în corpus cât mai multe tipuri de literatură posibile. Volumul total al corpusului este de 8,8 milioane de cuvinte; dicționarul de frecvențe format în baza corpusului acesta are vocabularul din mai mult de 200 000 de cuvinte. Este bine cunoscut faptul că cele mai frecvente cuvinte din text sunt așa-numite „stop-words”: articole, prepoziții, conjuncții, pronume și altele care nu dețin multă informație semantică și care sunt utilizate pentru formarea propozițiilor corecte din punct de vedere sintactic. Evident, nu am fost interesați în aceste cuvinte; am selectat cele mai frecvente 50 de substantive, 50 de adjective și 50 de verbe. Această listă de 150 de cuvinte aranjate în prima coloană a unui tabel a fost prezentată respondenților. Ei au trebuit să scrie în a doua coloană a tabelului

cuvântul pe care îl asociază în mintea lor în timp ce citează cuvântul din prima coloană a tabelului.

Respondenții au fost 50 de studenți cu vârste cuprinse între 19-21 de ani. Fiecare dintre ei a primit un document MSWord cu tabelul descris mai sus și au completat a doua coloană a tabelului. Am fost interesați de rezultatele statistice și anchetele au fost anonime.

Tabelul 1. Cele mai puternice asocieri obținute la prima etapă de creare a dicționarului asociativ român.

Cuvântul stimulent	Asocierea	Numărul de respondenți ce au oferit această asociere
forță	putere	29
ciudat	straniu	22
ceas	timp	21
noapte	întuneric	21
bucurie	fericire	18
istorie	trecut	18
târziu	noapte	18
moment	clipă	17
nevoie	necesitate	17
bucătărie	mâncare	15
frig	iarnă	15
piatră	tare	15

Datele obținute au fost grupate cu scopul de a găsi cele mai frecvente asocieri pentru fiecare cuvânt; am calculat numărul de câte ori a fost scrisă aceeași asociație pentru un cuvânt. De exemplu, pentru cuvântul „bucurie” 18 din 50 de respondenți au indicat „fericire”, 7 respondenți au indicat „zâmbet”, 6 respondenți au indicat „veselie”, celelalte asociații au fost diferite și au avut frecvență mai mică decât 3. Astfel cele mai puternice asociații pentru cuvântul „bucurie” au fost „fericire”, „zâmbet” și „veselie”. Noi am păstrat toate asociațiile oferite, chiar și cele cu frecvența egală cu 1, având în minte scopul de a mări dicționarul nostru asociativ.

Tabelul 1 conține cele mai frecvente 12 perechi de cuvinte stimulente și cuvinte asociate. De exemplu, perechea „forță – putere” are cea mai mare frecvență: 29 respondenți au indicat această asociere. Din punct de vedere a tipurilor asocierilor, cel mai frecvent au fost propuse sinonime sau cuvintele cu sensul similar. De exemplu, „ciudat – straniu”, „bucurie – fericire”, „moment – clipă”. Chiar dacă asocierea nu a fost sinonimă, ca în

³ <http://w3.usf.edu/FreeAssociation>

⁴ <http://www.eat.rl.ac.uk/>

⁵ <http://dictaverf.nsu.ru/fr>

⁶ <http://wordassociation.org>

⁷ <http://thesaurus.ru/dict/dict.php>

exemplul „bucătărie – mâncare”, în cele mai multe cazuri ea este de aceeași parte de vorbire ca și cuvântul de stimulare. Există, însă, un număr mic de excepții, de exemplu, „piatra – tare”.

Astfel, prima versiune a bazei de date asociative a fost obținută în urma prelucrării chestionarelor care conțin aproape 7 500 de perechi de stimulente-răspunsuri. A trebuit să eliminăm o parte din răspunsuri din diferite motive. Unii respondenți nu au fost atenți și au omis câteva cuvinte, unii au scris fraze lungi în loc de cuvinte ca răspunsuri, care nu au fost acceptat. După preprocesare am obținut 5821 de perechi diferite; 4152 perechi ce au avut frecvența egală cu 1.

3. A DOUA ETAPĂ A CREĂRII BAZEI DE DATE A ASOCIERILOR DE CUVINTE ROMÂNEȘTI

După prima etapă a creării dicționarului am obținut câte 50 de răspunsuri pentru fiecare din cele 150 cuvinte-stimulente selectate inițial. Această informație a fost organizată în baza de date MySQL, pe care ne-am propus s-o lărgim. În scopul de a obține mai multe cuvinte-asociații am creat o interfață online pentru dicționarul nostru folosind PHP⁸. Interfața este prezentată în figura 1. Aceasta poate fi accesată și interogată pe adresa <http://lilu.fcim.utm.md/asociere>.

Figura 1. Interfața pentru interogare a Dicționarului Asociativ Român.

După cum se vede în figura 1, utilizatorul poate introduce un cuvânt în casetă și apăsând butonul „Asocieri” obține toate perechile „stimulent

– asocierea” în care se întâlnește cuvântul dat. Figura 2 demonstrează rezultatul interogării pentru cuvântul „frumos”. Lista tuturor prechilor este sortată inițial conform frecvenței în ordine descrescătoare; ea poate fi de asemenea sortată în funcție de orice altă coloană a tabelului, în ordine descrescătoare sau crescătoare apăsând pe titlul coloanei respective cu săgeata corespunzătoare.

Există două tipuri de relații dintre cuvinte în dicționarul asociativ: relația directă de la stimulent spre răspuns, și relația inversă de la răspuns la stimulent; aceste relații nu sunt simetrice. Astfel, pentru stimulentele „aur” trei răspunsuri au fost „frumos”, iar în cazul în care „frumos” este stimulent niciun răspuns nu a fost „aur”.

#	Asociere ↑ / Asociere ↓	Asociere ↓ / Asociere ↑	Frecvența ↑ / Frecvența ↓ (Sortat.)
1	zâmbet	frumos	5
2	femeie	frumos	4
3	frumos	placut	4
4	viitor	frumos	4
5	aur	frumos	3
6	corpul	frumos	3
7	frumos	atrăgător	3

Figura 2. Asocierile pentru cuvântul „frumos” extrase din Dicționarul Asociativ Român.

Tabelul rezultat pentru interogare conține ambele tipuri de relații pentru cuvântul introdus ce se poate vedea în figura 2. Prima coloană conține cuvinte-stimulente; a doua conține cuvinte-răspunsuri. Cuvântul „frumos” apare în ambele coloane; în prima coloană ca stimulent și în a doua ca răspuns.

Ultima linie de text în interfața prezentată în figura 1 conține link-ul către pagina creată pentru introducerea de noi înregistrări în dicționarul asociativ. Această pagină este prezentată în figura 3.

Figura 3. Interfața pentru lărgirea Dicționarului Asociativ Român.

Un cuvânt aleatoriu este prezentat utilizatorului, iar utilizatorul trebuie să introducă cuvântul asociat în caseta de intrare. După ce face clic pe butonul „Asociază”, utilizatorul este

⁸ Baza de date și interfața au fost implementate de către fostul nostru student Ion Badan.

informat că asocierea introdusă a fost adăugată în baza de date. De exemplu:

„Baza de date a fost actualizată cu succes pentru înregistrarea lemn <-> foc”

Cu scopul de a extinde dicționarul nostru noi am realizat un algoritm de vizualizare a cuvintelor – stimulenți în mod aleatoriu. Inițial se formează lista de cuvinte din baza de date, care include atât stimulenții, cât și răspunsurile. Apoi din lista dată aleatoriu se selectează cuvintele pentru a fi afișate în calitate de stimulenți utilizatorilor online. Astfel, numărul de stimulenți de asemenea crește, depășind aceste 150 de cuvinte selectate inițial.

Tabelul 2. Cele mai puternice asocieri obținute pentru cuvântul „ceas”.

Cuvântul stimulent	Asocierea	Numărul de respondenți ce au oferit această asociere
ceas	timp	21
ceas	ora	16
timp	ceas	4
ceas	ac	2
ceas	trecere	2
ceas	frumos	2
ceas	vreme	2

Tabelul 3. Cele mai puternice asocieri obținute pentru cuvântul „piatra”.

Cuvântul stimulent	Asocierea	Numărul de respondenți ce au oferit această asociere
piatra	tare	14
piatra	roca	6
piatra	greutate	3
drum	piatra	2
piatra	colțuroasă	2
piatra	munte	2
piatra	casa	2

La momentul dat în baza de date sunt înregistrate 5965 de înregistrări „stimulent – răspuns – număr”, unde „număr” este numărul de câte ori acest răspuns a fost dat la stimulentele respective. În total sunt 9633 răspunsuri la cuvinte stimulente. De exemplu, perechea „forța – putere” are deja numărul 31, în comparație cu numărul 29 introdus inițial în baza de date, ce înseamnă că cuvântul acesta a fost propus utilizatorilor online ca

cuvânt-stimulent și în două cazuri a obținut același răspuns – „putere”.

Ca regulă, pentru fiecare cuvânt sunt înregistrate 2-3 relații puternice, introduse de un număr mai mare de persoane interogate și un număr mare de relații originale cu frecvența 1, introduse de o persoană. De exemplu, Tabelul 2 conține cele mai frecvente relații pentru cuvântul „ceas”, din care vedem că cele mai puternice sunt relațiile „ceas – timp” și „ceas – ora”, însă în afară de aceste relații în baza de date sunt stocate 20 de perechi cu frecvența 1 ca, de exemplu, „ceas – clipa”, „ceas – moment”, „ceas – orologiu”, „ceas – telefon”, etc. Alt exemplu este prezentat în Tabelul 3, asocierile frecvente pentru cuvântul „piatra”. Aici la fel observăm trei relații puternice „piatra – tare”, „piatra – roca”, „piatra – greutate”, însă mai există 10 perechi cu frecvența 2 și 24 perechi cu frecvența 1 ca de exemplu, „piatra – rece”, „piatra – funerară”, „piatra – prețioasă”, „piatra – nemișcată”, etc. Este interesant că pentru cuvântul dat au fost propuse mai multe adjective ce se combină cu substantivul acesta.

Trebuie de menționat că există și un număr de probleme care rămân a fi rezolvate. În primul rând, cuvintele adăugate online trebuie de verificat. Un utilizator poate adăuga informații greșite, un cuvânt cu erori gramaticale sau chiar o combinație de litere, fără nici un sens. Verificarea automată a unui dicționar poate înlătura cuvintele care nu sunt în dicționarul nostru și dacă cuvântul este scris cu o eroare gramaticală, este dificil să-l corectăm în mod automat. O parte din problema dată o reprezintă semnele diacritice. Unii utilizatori introduc cuvintele cu aceste semne; unii le ignoră, deoarece este o practică obișnuită în timp ce scriem online. Același cuvânt tastat în două forme, cu semne diacritice și fără ele, este considerat ca fiind două cuvinte diferite în baza de date. De exemplu, cuvântul stimulent „zice” are trei variante de răspuns: „vobeste”, „vorbeste” și „vorbește”. Primul dintre ele are o literă pierdută și nu are diacritice, al doilea este corect, dar fără diacritice și al treilea este absolut corect. Toate trei sunt stocate ca trei răspunsuri diferite în versiunea curentă a bazei de date. Astfel, intenționăm să găsim un lingvist pentru verificarea și corectarea înregistrărilor pe parcursul completării bazei de date.

4. CONCLUZII

Lucrarea raportează despre experimentul de creare a unui dicționar asociativ pentru limba română și anume despre a doua etapă ce include

crearea interfeței online și extinderea dicționarului prin internet. Sunt discutate o serie de aspecte tehnice ale etapei a doua. Sunt prezentate statistici preliminare și se face o scurtă analiză a bazei de date obținute. Dicționarul creat poate fi utilizat în filologie, lexicografie și studierea limbii române. La etapa această de lucru noi însă suntem interesați în crearea unei baze de date mai bogate și mai reprezentative de asocieri de cuvinte; analiza detaliată este amânată pentru cercetarea viitoare.

Bibliografie

1. **Miller George A.** *WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41, 1995.*
2. **Vossen P.** Ed. 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks.* Kluwer, Dordrecht, The Netherlands.
3. **Alessandro Lenci, Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, and Antonio Zampolli.** *SIMPLE: A General Framework for the Development of Multilingual Lexicons. International Journal of Lexicography, 13(4):249–263, 2000.*
4. **Catherine Havasi, Rob Speer and Jason Alonso.** *ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge. Proceedings of Recent Advances in Natural Language Processing, 2007.*
5. **Navigli R., Ponzetto S.** *BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. Artificial Intelligence, 193, Elsevier, 2012, pp. 217-250.*
6. **Nelson D. L., McEvoy C. L., & Schreiber T. A.** *The University of South Florida word association, rhyme, and word fragment norms.* <http://www.usf.edu/FreeAssociation/>, 1998.
7. **Pustejovsky J.** (2010) *Qualia Roles. The Cambridge Encyclopedia of the Language Sciences.* Ed. Patrick Hogan. Cambridge, UK: Cambridge University Press.
8. **Witbrock M., Baxter D., Curtis J.** 2003. *An Interactive Dialogue System for Knowledge Acquisition in Cyc. Eighteenth International Joint Conference on Artificial Intelligence.*
9. **Cramer P.** *Word association.* New York & London: Academic Press, 1968.
10. **Deese J.** *The structure of associations in language and thought.* Baltimore: The John Hopkins Press, 1965.
11. **Christopher D. Manning, Hinrich Schütze.** *Foundations of Statistical Natural Language Processing.* 1999, The MIT Press, Cambridge, Massachusetts.
12. **Hirst G.** (2004). *Ontology and the lexicon.* In Steffen Staab, & Rudi Studer, (Eds.), *Handbook of ontologies.* (pp. 209-229). Berlin, Heidelberg, & New York: Springer-Verlag.
13. **Collins A. M., Loftus E. F.** *A spreading-activation theory of semantic processing. Psychological Review, 82, 407-428, 1975.*
14. **Steyvers M., Shiffrin R. M., Nelson D. L.** (2004). *Word association spaces for predicting semantic similarity effects in episodic memory.* In A. F. Healy, (Ed.), *Experimental cognitive psychology and its applications. (Decade of behavior).* (pp. 237-249). Washington, D.C.: American Psychological Association.
15. **Kiss G.R., Armstrong C., Milroy R., Piper J.** *An associative thesaurus of English and its computer analysis.* In Aitken, A.J., Bailey, R.W. and Hamilton-Smith, N. (Eds.), *The Computer and Literary Studies.* Edinburgh: Edinburgh University Press, 1973.
16. **Karaulov Iu. N., Cercasova G. A., Ufimțeva N. V., Sorochin Iu. A. Iaroshinscaia V. N.,** *Russkij Assotziativnyj slovar' . Tom I, Ot stimula k reactzii. Tom II, Ot reactzi k stimulu. Astreli. AST, 784 (992) pag. 2002,2003.*
17. **Baltova P., Eftimova A., Lipovska A., Petrova K.** *BAS 2003: Bolgarski asoțiativen rechnik, Sofia, Izd. SU "Sv. Kl. Ohridski", 2003.*
18. **Ufimțeva N. V.** *Slavianskii asoțiativnii slovari: ruskii, beloruskii, bolgarskii, ukrainskii. Institut iazikoznaniia RAN, 790 pag. 2004.*
19. **Joyce Terry.** *Constructing a large-scale database of Japanese word associations. (Special issue edited by Katsuo Tamaoka: Corpus Studies on Japanese Kanji). Glottometrics, 10, 82-98, 2005.*
20. **Vlad A., Mitrea A., Mitrea M.** *Limba română scrisă ca sursă de informație. Paideia, România. 2005.*

Recomandat spre publicare: 29.10.2015.