

## EMOȚII ÎN CUVINTE: ELABORAREA RESURSEI MULTILINGVE

VICTORIA BOBICEV, VICTORIA MAXIM, TATIANA PRODAN,  
NATALIA BURCIU, VICTORIA ANGHELUȘ

*Universitatea Tehnică a Moldovei, Chișinău, Moldova*

*vika@rol.md, maxivica@yahoo.com, tatiana.ursulenco@gmail.com,  
natusicb@yahoo.com, lazu\_vic@yahoo.com*

### Rezumat

În lucrarea dată este descris procesul de creare a WordNet-Affect pentru limbile română și rusă. WordNet-Affect reprezintă o resursă lexicală elaborată în baza WordNet-ului englez ce conține informații despre emoțiile pe care le transmit cuvintele. WordNet-Affect este organizat în șase emoții de bază: *anger, disgust, fear, joy, sadness, surprise*.

Noi am tradus cuvintele din WordNet-Affect în limbile română și rusă și am creat o resursă lexicală aliniată engleză - română - rusă. Resursa dată este accesibilă gratuit în scopuri de cercetare.

### 1. Introducere

Creșterea spectaculoasă a tehnologiilor Web 2.0 permite fiecărui utilizator să participe activ la crearea de conținut web (bloguri, rețele sociale, chat-uri). Volumul de texte cu conținut emoțional bogat crește în progresie geometrică. Acest lucru face ca analiza subiectivă a textelor să fie de actualitate în mod special.

Până în prezent, analiza sentimentelor și studiile asupra afectelor cuvintelor s-au concentrat pe limba engleză. Un exemplu este sarcina SemEval-2007 de clasificare a textelor afective (Liu et al., 2003). De asemenea, cele mai multe resurse lexicale au fost create pentru limba engleză. De exemplu, SentiWordNet este o resursă lexicală pentru extragerea opiniilor care atribuie fiecărui synset al Wordnet-ului trei marcaje ale sentimentelor: pozitivitate, negativitate, obiectivitate (Esuli, Sebastiani, 2006).

Recent, cea mai mare creștere a utilizării Internetului a fost înregistrată pentru vorbitorii de limbi diferite de limba engleză: în anii 2000-2009, pentru limbile diferite de engleză, creșterea a depășit 690% în comparație cu 237% a creșterii pentru limba engleză<sup>1</sup>.

În consecință, cantitatea informației textuale scrise în alte limbi decât engleza crește rapid ce sporește cererea pentru instrumentele de analiză automată a textelor și pentru resursele lexicale pentru alte limbi, diferite de engleză. Elaborarea instrumentelor a progresat pentru limbile vest-europene (franceză, germană) și asiatice (japoneză, chineză, arabă) (Edmonds, 2004), se simte nevoia de astfel de resurse și pentru limbile est-europene. Pentru a umple acest gol, am elaborat o resursă lexicală, în baza WordNet-Affect, prin traducerea în română și rusă, redactarea synset-urilor traduse și alinierea acestora la sursa engleză.

<sup>1</sup> <http://www.internetworldstats.com/stats7.htm>

## 2. WordNet-Affect

WordNet-Affect<sup>2</sup> este o resursă lexicală ce conține informații despre emoțiile pe care le transmit cuvintele. În comparație cu WordNet-ul complet, WordNet-Affect este o resursă lexicală mică, dar valoroasă pentru adnotarea sa afectivă.

a#01943022 awed awestruck awestricken in\_awe\_of

**Figura 1:** Un synset din WordNet-Affect.

WordNet-Affect (Strapparava, Valitutti, 2004) a fost creat pornind de la WordNet DOMAINS (Magnini, Cavaglia, 2002). WordNet-Affect produce o ierarhie suplimentară a etichetelor domeniilor afective, independent de ierarhia domeniilor, cu care sunt adnotate synset-urile ce reprezintă concepte afective. "Cuvintele afective" sunt considerate a fi cuvintele ce au "conotație emoțională" (Ortony, 1987). Există cuvinte ce nu descriu direct unele emoții (de exemplu, bucurie, tristețe sau frică), dar, de asemenea, sunt legate de emoții precum cuvintele ce descriu stările mintale, stările fizice sau trupești, trăsăturile de personalitate, comportamente, atitudini și sentimente (cum ar fi plăcere sau durere).

Tabel 1: Seturile de date ale cuvintelor afective.

Clasele	#synset-uri	%synset-uri	#cuvinte	%cuvinte
<b>anger</b>	128	21.0	318	20.7
<b>disgust</b>	20	3.3	72	4.7
<b>fear</b>	83	13.5	208	13.5
<b>joy</b>	228	37.2	539	35.1
<b>sadness</b>	124	20.3	309	20.1
<b>surprise</b>	29	4.7	90	5.9
<b>Total</b>	612	100.0	1536	100.0

Colecția de synset-uri din WordNet-Affect utilizată în lucrarea dată a fost furnizată ca o resursă pentru SemEval-2007 "Affective Text"<sup>3</sup>. Această sarcină a fost axată pe adnotarea textelor prin etichete afective (Strapparava, Mihalcea, 2008). Toate synset-urile din WordNet-Affect sunt adnotate utilizând șase etichete de categorii emoționale: *joy*, *fear*, *anger*, *sadness*, *disgust*, *surprise* (Strapparava et al., 2006). Această alegere a celor șase emoții rezultă din cercetarea psihologică a emoțiilor umane exprimate non-verbal (Ekman, 1992).

Datele despre resursa inițială sunt descrise în tabelul 1. Toată resursa este prezentată în șase fișiere denumite conform celor șase emoții. Fiecare fișier conține o listă de synset-uri; un synset pe un singur rând. Un exemplu de synset este prezentat în figura 1.

Prima literă din rând indică partea de vorbire; aceasta este urmată de numărul synset-ului și apoi sunt enumerate toate cuvintele din synset. Această reprezentare a fost simplă și ușoară pentru prelucrarea ulterioară. În synset-uri inițiale este un număr mare de combinații de cuvinte, sintagme și expresii. Un exemplu poate fi văzut în figura 1: "in\_awe\_of". Aceste părți ale synset-urilor au generat probleme în timpul traducerii.

<sup>2</sup> În scopuri de cercetare, WordNet-Affect este accesibil la cerere la <http://wndomains.itc.it>

<sup>3</sup> <http://www.cse.unt.edu/~rada/affectivetext>

### 3. Elaborarea WordNet-ului pentru limbile română și rusă

WordNet-ul Român a fost creat de către Universitatea Alexandru Ioan Cuza din Iași în cadrul proiectului european Balkanet (Tufiş et al., 2006). După ce proiectul Balkanet s-a încheiat, Institutul de Cercetări pentru Inteligență Artificială, de la Academia Română a continuat să lucreze cu WordNet-ul Român și, în prezent, acesta conține 33151 synset-uri de substantive, 8929 synset-uri de verbe, 851 synset-uri de adjective și 834 synset-uri de adverbe (Tufiş et al., 2008). Acesta poate fi accesat prin intermediul interfeței online a MultiWordNet<sup>4</sup> în care WordNet-urile pentru câteva limbi sunt aliniate la WordNet-ul Princeton.

În lucrarea dată, mai întâi de toate am verificat synset-urile din WordNet-Affect folosind interfața online a MultiWordNet. Astfel, am copiat la setul nostru toate synset-urile care deja există în WordNet-ul Român și nu am procesat aceste synset-uri suplimentar. Ca rezultat, în WordNet-ul Român au fost găsite 166 synset-uri, majoritatea dintre ele fiind pentru substantive și verbe. Adjectivele și adverbele sunt reprezentate mai puțin. Statistica privind synset-urile deja existente în română este prezentată în tabelul 2.

Tabel 2: Seturile de date a synset-urilor deja existente în Wordnet-ul Român.

Clasele	# synset-uri în WordNet-Affect	# synset-uri din WordNet Român	% synset-uri din WordNet Român
<b>anger</b>	116	35	30.1
<b>disgust</b>	17	7	41.1
<b>fear</b>	76	25	32.8
<b>joy</b>	209	63	30.0
<b>sadness</b>	98	24	24.7
<b>surprise</b>	26	12	46.1
<b>Total</b>	542	166	30.6

Pentru limba rusă situația este complet diferită. Au fost făcute mai multe încercări pentru a crea WordNet-ul pentru limba rusă. RussNet este un proiect pentru crearea tezaurului rusesc electronic (Azarova et al., 2002). Un proiect alternativ al versiunii rusești a WordNet-ului este WordNet-ul Rus (Balkova et al., 2004). Ambele proiecte sunt necomerciale. De asemenea, există două proiecte comerciale menite să elaboreze WordNet-uri pentru limba rusă: RuThes este un tezaur informațional utilizat în UIS RUSSIA<sup>5</sup> și proiectul privind WordNet-ul Rus creat de către grupul companiei Novosoft<sup>6</sup>. Din păcate, puțină informație și încă mai puține resurse sunt accesibile în mod gratuit.

### 4. Crearea WordNet–Affect pentru limbile română și rusă

Pentru crearea acestor două seturi de date, am parcurs trei etape: (1) traducere automată; (2) înlăturarea traducerilor irelevante; (3) generarea synset-urilor pentru română și rusă.

<sup>4</sup> <http://multiwordnet.itc.it/english/home.php>

<sup>5</sup> <http://www.cir.ru>

<sup>6</sup> <http://research-and-development.novosoft-us.com>

#### 4.1 Traducerea Automată

Traducerea a fost făcută automat utilizând dicționare bilingve. Am folosit dicționarul electronic Română–Englez ROMEN de la PRIMASOFT<sup>7</sup>. Dicționarul constă din următoarele compartimente: Englez–Română, Română–Englez, Englez–Rus, Rus–Englez, fiecare conținând mai mult de 200 000 de intrări. Noi am utilizat doar compartimentele în care limba engleză este limba sursă. În dicționare au fost combinații de cuvinte, sintagme și expresii pe care le-am folosit în limbile țintă. Pentru traducerea automată, dicționarul a fost organizat într-o listă de cuvinte sursă urmate de traducerile țintă. Un exemplu de intrare a dicționarului este prezentat în figura 2.

```
Joy  
Dicționar general:  
noun: bucurie; confort; fericire; plăcere; tihnă; veselie;  
voioșie;  
verb: a bucura; a înveseli;
```

**Figura 2:** Un exemplu de intrare a dicționarului.

La această etapă, scopul nostru a fost să obținem cât mai multe cuvinte afective posibile pentru analiză. Pentru aceasta, am tradus fiecare cuvânt din synset-urile WordNet-Affect. Am decis să excludem din synset-urile engleze toate combinațiile de cuvinte, sintagmele și expresiile, deoarece ele nu pot fi traduse automat. Figura 3 prezintă un exemplu de synset tradus, obținut după această etapă. După cum se vede în exemplu, pentru traducerea română, de asemenea am obținut combinații de cuvinte care erau în dicționar: ”a avea gust”, ”a degusta (un aliment)”.

Unele elemente ale synset-urilor nu au fost traduse. Acestea pot fi divizate în patru grupe. (1) Combinații de cuvinte, sintagme și expresii pe care le-am înlăturat intenționat din synset-urile engleze înainte de traducere. (2) Variații în ortografia aceluiași cuvânt; de exemplu, ”jubilation”, ”jubilationcy” – primul cuvânt a fost tradus, al doilea nu a fost găsit în dicționar. (3) Cuvinte care au fost formate cu ajutorul sufixelor ”ness”, ”less”, ”ful” (de exemplu ”heartlessness”); este puțin probabil ca acestea să apară în dicționare, la fel ca și adverbele formate cu sufixul ”ly”. (4) Cuvinte care nu au fost traduse din cauza limitărilor dicționarului utilizat. În timp ce WordNet poate fi, în mod rezonabil, menționat ca unul din cele mai mari dicționare engleze, dicționarul nostru bilingv este destul de modest. Tabelul 3 arată procentajul cuvintelor care nu au fost traduse. Media în procente a cuvintelor netraduse a fost 21%.

---

<sup>7</sup> [http://www.primasoft.biz/romen\\_eng.php](http://www.primasoft.biz/romen_eng.php)

## EMOȚII ÎN CUVINTE: ELABORAREA RESURSEI MULTILINGVE

Tabel 3: Numărul și procentajul cuvintelor netraduse.

Clase	# de cuvinte engleze	# de cuvinte traduse	# de cuvinte netraduse	% de cuvinte netraduse
<b>anger</b>	318	248	70	22.0
<b>disgust</b>	72	60	13	18.0
<b>fear</b>	208	162	47	22.5
<b>joy</b>	539	420	119	22.0
<b>sadness</b>	309	246	63	20,5
<b>surprise</b>	90	72	18	21.0
<b>Total</b>	<b>1536</b>	<b>1208</b>	<b>330</b>	<b>21.0</b>

A doua grupă de cuvinte nu prezintă o problemă, dar prima, a treia și a patra au trebuit traduse manual. Aceasta s-a făcut la etapa a treia.

```

05573914 n:
preference =
    preferință
penchant =
    înclinație
    slăbiciune
predilection =
    predilecție
taste =
    a avea gust
    a gusta, a cunoaște
    a gusta; a degusta (un aliment)
    degustare
    fărămă, bucățică, îmbucătură (de)
    gust
    înclinație, preferință
    
```

**Figura 3:** Un exemplu de traducere a unui synset.

### 4.2 Înlăturarea Traducerilor Irelevante

Multe cuvinte din synset-urile engleze au câteva sensuri. Este evident că traducerea automată a generat toate traducerile posibile pentru toate sensurile. Noi am ales doar o singură traducere care era relevantă pentru semnificația synset-ului. Traducerea relevantă a fost selectată manual. Am înlăturat toate traducerile a căror sensuri nu aveau legătură cu emoția. De exemplu, cuvântul "taste" în synset-ul cu înțelesul de "preference" are câteva semnificații, însă doar ultima din lista traducerilor posibile avea legătură cu sensul comun al synset-ului. Exemplul este prezentat în figura 3. Astfel, noi am înlăturat toate traducerile cu excepția ultimei.

Deoarece am tradus fiecare cuvânt în mod separat, am obținut o mulțime de duplicate care la fel au trebuit eliminate. Am tras atenția, de asemenea, la corespondența părților de vorbire. În multe cazuri, a fost destul de dificil, în special pentru substantivele deja menționate, formate cu ajutorul sufixelor, de exemplu, "plaintiveness" sau "uncheerfulness".

### 4.3 Generarea synset-urilor române și ruse

Toate cuvintele dintr-un synset reprezintă un concept, un singur sens. Scopul etapei a treia a fost de a găsi traducerea adecvată, exact corespunzătoare acestui sens. La această etapă, a trebuit mai întâi să atașăm explicația în engleză a fiecărui synset. Aceasta a determinat claritatea în semnificația synset-ului pentru traducători. După ce au fost adăugate explicațiile la synset-uri, întregul set a fost dat la trei traducători care au lucrat independent. Sarcina lor a fost dublă: (1) să elimine traducerile care, din punctul lor de vedere, nu erau relevante semnificației synset-ului descrise de explicația acestuia; (2) să adauge cât mai multe sinonime posibile relevante la synset-urile române și ruse. Astfel, sarcina lor a fost să verifice echivalența sensurilor synset-urilor engleze, ruse și române. De asemenea, au trebuit să traducă cuvintele care au rămas netraduse la prima etapă. Pentru traducere au fost folosite dicționare online.

Dicționarele bilingve române utilizate:

- <http://hallo.ro>,
- <http://dictionar.netflash.ro>,
- <http://www.ectaco.co.uk/English-Romanian-Dictionary>;

Dicționar explicativ român: <http://dexonline.ro/>.

Dicționarele bilingve ruse utilizate:

- <http://en.bab.la>,
- <http://dictionary.babylon.com>,
- <http://russianlessons.net/dictionary/dictionary.php>;

Dicționare explicative ruse:

- <http://slovo.freecopy.ru/>,
- <http://slovari.yandex.ru/dict/ushakov>.

Această etapă a fost cea mai dificilă care a necesitat cel mai mult lucru. Multe synset-uri engleze aveau aproape aceleași sensuri, doar cu careva nuanțe. În unele cazuri, synset-urile conțineau cuvinte arhaice, care nu erau găsite în dicționare. După cum s-a menționat mai sus, am încercat să evităm combinațiile de cuvinte, sintagmele și expresiile. Cu toate acestea, în unele cazuri, sensul exact al unui synset englez putea fi reprezentat doar de o singură combinație de cuvinte române sau ruse. În unele cazuri, chiar și synset-ul englez era prezentat de combinații de cuvinte. De exemplu, n#05591681 stage\_fright. Un alt exemplu conține un cuvânt german: n#05600844 world-weariness Weltschmerz. În așa cazuri, nu am obținut traducerea corespunzătoare. În unele cazuri, câteva synset-uri engleze au fost traduse în aceleași cuvinte române sau ruse deoarece nu am putut reflecta în limba țintă nuanțele sensurilor limbii sursă.

Referindu-se la problema cu sufixe, de exemplu, cuvintele "weepiness", "plaintiveness", "mournfulness", "ruthfulness", cu greu pot fi găsite în dicționare. Pentru a rezolva această problemă, am eliminat sufixele și am căutat rădăcinile cuvintelor menționate în dicționare disponibile. În acest fel, am putut găsi sensul cuvintelor și, prin adăugarea afixelor necesare, au fost create echivalentele în

română și rusă. De exemplu, pentru a găsi traducerea adecvată pentru cuvântul "mournfulness", am căutat în dicționar cuvântul "mournful". Rezultatul pentru limba română este "îndoliat" și pentru rusă "траурный". Având în vedere că cuvântul "mournfulness" este un substantiv, am transformat adjectivele obținute în substantive. Astfel, echivalentul românesc este "doliu" și în rusă - "траур".

Cu toate acestea, cele mai multe probleme au apărut la alinierea adjectivelor. De exemplu, pentru eticheta emoțională "*sadness*", multe synset-uri adjectivale traduse în limba rusă conțin cuvintele "грустный" și "печальный". Prin urmare, pentru synset-uri adjectivale diferite obținem traduceri destul de asemănătoare.

#### 4.4 Acordul între traducători

Având ca rezultatul traducerii seturi de sinonime, nu am putut folosi măsurile standard pentru acordul între traducători. Astfel, acordul a fost calculat după cum urmează. Dacă A este o mulțime de cuvinte selectate de către primul traducător pentru un synset și B este o mulțime de cuvinte selectate de al doilea traducător pentru același synset, acordul între traducători  $IntAgr$  este egal cu raportul dintre numărul de cuvinte în intersecția mulțimilor A și B și numărul de cuvinte în uniunea mulțimilor A și B:

$$IntAgr = (A \cap B) / (A \cup B) \quad (1)$$

De exemplu, dacă un traducător a format un synset din trei cuvinte  $w_k, w_l$  și  $w_m$ , iar al doilea traducător a format același synset din patru cuvinte  $w_k, w_l, w_m$  și  $w_n$ , și primele trei cuvinte sunt aceleași, atunci  $A = (w_k \ w_l \ w_m)$ ,  $B = (w_k \ w_l \ w_m \ w_n)$ ,  $A \cap B = (w_k \ w_l \ w_m)$ ,  $A \cup B = (w_k \ w_l \ w_m \ w_n)$ , numărul de cuvinte din intersecția A și B ar fi egal cu 3, numărul de cuvinte din uniunea A și B ar fi egal cu 4 și, prin urmare, acordul între traducători - ar fi  $3 / 4 = 0.75$ .

Spre exemplu synset-ul "a#01195320 friendly" a fost tradus de primul traducător ca "prietenos prietenesc amical", de al doilea traducător ca "binevoitor amical prietenos", și de al treilea, ca "binevoitor prietenesc prietenos". Pentru primul și al doilea traducător intersecția traducerilor era de două cuvinte: "prietenos amical" și uniunea traducerii a fost patru cuvinte "prietenos prietenesc binevoitor amical". Acordul între traducători, în acest caz a fost de  $2 / 4 = 0.5$ . Pentru al doilea și al treilea traducător, intersecția traducerilor era de două cuvinte: "prietenos binevoitor", și uniunea - patru cuvinte "prietenos prietenesc binevoitor amical". Prin urmare, acordul este același: 0.5. Pentru primul și al treilea traducător acordul, de asemenea, este același: 0.5. Toți cei trei traducători au în comun doar un singur cuvânt "prietenos" și uniunea de traduceri a constat din patru cuvinte. Astfel, acordul a fost de  $1 / 4 = 0.25$ .

Tabelul 4 prezintă valorile medii ale acordului între traducători. Cei trei traducători sunt prezentați ca T1, T2 și T3.

Tabel 4: Acordul între traducători.

Perechile de traducători	Acordul între traducători
<b>Date pentru limba rusă</b>	
T1 – T2	0.57
T2 – T3	0.61
T1 – T3	0.59
Toți	0.29
<b>Date pentru limba română</b>	
T1 – T2	0.58
T2 – T3	0.57
T1 – T3	0.67
Toți	0.32

Unele synset-uri aveau acordul egal cu unu, spre exemplu, în synset-ul "a#00863650 euphoriant", toți trei traducători au tradus cuvânt respectiv ca "euforizant". Cu toate acestea, pentru majoritatea synset-urilor, traducătorii au furnizat mai multe traduceri diferite, dar nu multe dintre aceste traduceri au fost comune pentru toți traducătorii. În unele synset-uri traduse, nu a fost găsit nici un cuvânt comun în traducerile celor trei traducători. De exemplu, pentru synset-ul "a#00670851 gladdened exhilarated", cele trei traduceri au fost "bucurat înveselit înviorat bine\_dispus", "bucuros vesel voios încântat bine\_dispus" și "bucurat voios bucuos înveselit". Nu a fost găsit nici un cuvânt comun pentru toate cele trei traduceri. Astfel, ne-am decis să formăm synset-uri din cuvintele care au apărut în cel puțin două dintre cele trei variante de traduceri. În acest fel, am format synset-uri finale. De exemplu, synset-ul "a#01195320 friendly" a fost tradus ca "prietenos prietenesc binevoitor amical", pentru că toate aceste cuvinte au apărut de cel puțin două ori în traduceri. Synset-ul "a#00670851 gladdened exhilarated" a fost tradus ca "bucurat înveselit bine\_dispus bucuos voios".

Tabelul 5 conține date privind numărul final de cuvinte în traduceri pentru fiecare din cele șase emoții din WordNet-Affect .

Tabel 5: Seturile de date a cuvintelor afective din limba română și rusă.

Clasele	#synset-urile	# Cuvinte ruse	# Cuvinte române
<b>anger</b>	116	393	330
<b>disgust</b>	17	73	60
<b>fear</b>	76	327	248
<b>joy</b>	209	765	641
<b>sadness</b>	98	437	364
<b>surprise</b>	26	129	87
<b>Total</b>	542	2199	1869

Trebuie de menționat că în sursa WordNet-Affect englez au fost găsite unele synset-uri duplicate. Am eliminat toate aceste repetiții și numărul de synset-uri în sursa noastră este mai mic. În plus, au fost depistate unele diferențe mici dintre WordNet-Affect,



MultiWordNet și versiunea online a WordNet pentru că WordNet-Affect s-a creat în baza WordNet versiunea 1.6, MultiWordNet utilizează versiunea 2.0 a WordNet-ului și versiunea online a WordNet-ului este 3.0. În ciuda numărului mai mic de synset-uri, numărul de cuvinte în seturile pentru limbile română și rusă este mai mare decât în limba engleză. Acest lucru se datorează tendinței noastre de a colecta în resursele noastre cât mai multe cuvinte posibile. Scopul nostru este de a folosi această resursă pentru metodele statistice de recunoaștere a emoțiilor în text.

### 5. Concluzie și planuri pe viitor

Acest articol descrie procesul de creare a resursei lexicale multilingve WordNet-Affect ce conține partea engleză, partea rusă și partea română aliniată la nivel de synset. WordNet-Affect este o resursă lexicală creată în baza WordNet-ului Princeton care codifică informațiile despre emoțiile pe care le transmit cuvintele. Acesta este organizat în șase emoții de bază: *anger, disgust, fear, joy, sadness, surprise*. WordNet-Affect este o resursă lexicală comparativ mică, dar valoroasă pentru adnotările sale afective.

În lucrare este descris procesul traducerii synset-urilor WordNet-Affect în limba română și rusă și creării WordNet-Affect aliniat Engleză - Română - Rusă. Resursa poate fi folosită pentru recunoașterea automată a emoțiilor și afectelor în text. Aceasta poate fi obținută gratuit în scopuri de cercetare, pe <http://lilu.fcim.utm.md>. Resursa este încă în dezvoltare. Prima versiune bazată pe WordNet-Affect a fost lansată în august 2009; a doua, lansată în octombrie 2009, este aliniată cu WordNet-ul român. Resursa a fost deja folosită în (Sokolova, Bobicev, 2009) și aceasta este numai una dintre multiplele posibilități de utilizare a seturilor de cuvinte.

În viitor plănuim să mărim resursa noastră. Recent am obținut WordNet Domains care este distribuit cu WordNet-Affect. În resursa dată sunt incluse două versiuni ale WordNet-Affect original englez. Prima versiune a fost obținută semiautomat și conține aproximativ 1 500 de synseturi. A doua versiune a fost creată după verificarea manuală a primei versiuni și excluderea synset-urilor care aveau o legătură mai slabă cu emoția respectivă. Versiunea aceasta conține 914 synset-uri. Intenția noastră este de a traduce toate 914 synset-uri și de a alinia acestea la WordNet român. În afară de aceasta, vom crea resurse 'bag-of-words' pentru utilizarea imediată în sisteme de recunoaștere a emoțiilor și afectelor în text.

### Referințe bibliografice

- Azarova, I., Mitrofanova, O., Sinopalnikova, A., Yavorskaya, M., Oparin I. (2002). Russnet: Building a lexical database for the russian language. *Workshop on Wordnet Structures and Standardization and How this affect Wordnet Applications and Evaluation*, Las Palmas, pp. 60-64.
- Balkova V., Suhonogov A., Yablonsky S.A. (2004). Russian WordNet. From UML-notation to Internet/Intranet Database Implementation. *Second International WordNet Conference, GWC 2004*, Brno, Czech Republic, 31-38.
- Edmonds, P. (2002). Introduction to Senseval. *ELRA Newsletters*, 7(3), 337-344.

- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, vol. 6(3-4), 169–200.
- Liu, H., Lieberman, H., Selker, T. (2003). A model of textual affect sensing using real-world knowledge. *ACM Conference on Intelligent User Interfaces*, Miami, Florida, USA, pp. 125-132.
- Strapparava, C., Mihalcea, R. (2008). Learning to identify emotions in text. *ACM Symposium on Applied Computing*, Fortaleza, Brazil, 556-1560.
- Strapparava, C., Valitutti, A. (2004). Wordnet-affect: an affective extension of wordnet. *4th International Conference on Language Resources and Evaluation*, 1083–1086.
- Strapparava, C., Valitutti, A., Stock, O. (2006). The affective weight of the lexicon. *5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 474–481.
- Tufiş, D., Mititelu, B., Bozianu, L., Mihaila, C. (2006). Romanian wordnet: New developments and applications. *3rd Conference of the Global WordNet Association*, Korea, 337–344.
- Tufiş, D., Ion, R., Bozianu, L., Ceauşu, A., Ştefănescu, D. (2008). Romanian Wordnet: Current State, New Applications and Prospects. *4th Global WordNet Conference, GWC-2008*, University of Szeged, Hungary, 441-452.
- Esuli, A., Sebastiani, F. (2006). SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. *5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 417-422.
- Magnini, B., Cavaglia G. (2002). Integrating subject field codes into Wordnet. *Second International Conference on Language Resources and Evaluation (LREC 2002)*, Athens, Greece, 1413—1418.
- Ortony, A., Clore, G. L., Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology*, American Psychological Association, vol. 53, 751–766.
- Sokolova M., Bobicev V. (2009). Classification of Emotion Words in Russian and Romanian Languages. *RANLP-2009 conference*, Borovets, Bulgaria, 415-419.