# A Cross-Language Linguistic Resource for Lexical Representation of Affects

**Victoria BOBICEV**
*Technical University of Moldova*
*vika@rol.md*

**Victoria MAXIM**
*Technical University of Moldova*
*maxivica@yahoo.com*

*Abstract*—This paper describes the process of Russian and Romanian WordNet-Affect creation. WordNet-Affect is a lexical resource created on the base of Princeton WordNet which contains information about emotions the words convey. It is organized in six basic emotions: anger, disgust, fear, joy, sadness, surprise. WordNet-Affect is a small lexical resource but valuable for its affective annotation.

We translated WordNet Affect synsets into Russian and Romanian and, then, create two data sets for learning experiments. The two data sets will be available for research purposes upon publication.

*Index Terms*—Natural Language Processing, Automatic Text Analysis, Eastern European languages, Sentiment Analysis, lexical representation of affects.

## I. INTRODUCTION

How some people are able to express emotion through words, while others can not? Why reading some text, you feel his emotion, while the other reading is striking you with slough of compact dead letters? Why someone can give life to the written words, while others kill them every time they use them? I have only one explanation: a double distinction between words and representations on the one hand, and representations and emotions on the other.

All words can transmit affective meaning. Every word, even those apparently neutral, can evoke pleasant or painful experiences due to their semantic connections with the emotional concepts or categories. Some words carry emotional meaning of a particular text, while for many others is the emotional part of the collective imagination.

The automatic detection of emotions in text is increasingly important in different applications. Emotional or affective calculation is developing as a field that allows a new form of human computer interaction, added to natural language use. There is a wide range of perception that future human-computer interaction consists of such topics as entertainment, emotions, aesthetic pleasure, motivation, attention, commitment, etc. Studying the link between natural language and affective information becomes increasingly crucial, having to deal with its computational treatment.

So far, sentiment analysis and studies of the word affect concentrated on English. Our aim is to create affective lexical resources for other languages based on the WordNet-Affect domains.

In this paper we describe the building of two data sets of affective words, in Russian and Romanian languages respectively.

## II. MOTIVATION

Recently, most of the Internet use growth was supported by non-native English speakers: starting 2000, for non-English speaking regions, the growth has surpassed 3,000 % to compare with 342 % of the over-all growth.[1]

Consequently, the amount of text data written in languages other than English rapidly increased [3]. This raise increased the demand for automated text analysis tools in languages other than English. The tool development has progressed for Western European (French, German) and Asian (Japanese, Chinese, Arabic) [4]. However, some languages did not attract as much attention of NLP and Text Data Mining community, e. g. Eastern European.

## III. WordNet, WordNet-Affect

WordNet[2] is a lexical database for the English language. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. The purpose is twofold: to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications. WordNet was created and is being maintained at the Cognitive Science Laboratory of Princeton University under the direction of psychology professor George A. Miller.

WordNet represents the largest publicly available on-line lexical resource, already used in various applications of the human language technology. Systems performing word sense disambiguation, information extraction or retrieval, prepositional attachment and many other natural language processing tasks use WordNet.

WordNet-Affect [8] is an additional hierarchy of "affective domain labels", with which the synsets representing affective concepts are further annotated.

WordNet-Affect[3] is a well-used lexical resource which contains information about emotions the words convey. It has been developed through a selection and labelling of the affective concepts represented by sets of synonyms,

---

[1] http://www.internetworldstats.com/stats.htm
[2] http://wordnet.princeton.edu/

[3] For research purposes, WordNet-Affect is available upon request at http://wndomains.itc.it

based on the lexical knowledge of the WordNet. If compared with the complete Word-Net, WordNet Affect is a small lexical resource but valuable for its affective annotation.

A number of affective labels (a-labels) were manually assigned to the WordNet synsets of nouns, adjectives, verbs and adverbs which convey affective meaning. Words that have an Emotion tag were lately more fine-grain reannotated using six emotional category labels: joy, fear, anger, sadness, disgust, surprise [9]. This choice of the six emotions comes from psychological research into human non-verbally expressed emotions [5].

The collection of the WordNet-Affect synsets used in our work is described in Table 1. The data was provided as a resource for the SemEval-2007 "Affective Text", the task focused on text annotation by affective tags [7]. The whole data is provided in six files named by the six emotions. Each file contains a list of synsets in following form:

```
a#01943022 awed awestruck
    awestricken in_awe_of
```

First letter in the line indicates part of speech; it is followed by number of the synset and then all synset words are listed. There are a large number of word combinations, collocations and idioms. One of them can be seen in the example.

### IV.       Applications of WordNet-Affect

WordNet-Affect is useful in all applications in which it is necessary to have an affective interaction [8].

Affective text sensing systems are the programs for assessing the affective qualities of natural language. An interesting approach, corpus-based, is that of [6]. The affect of the text, at the sentence level, is classified into one of six basic categories of emotions. The analysis is performed through a model built starting from OpenMind Commonsense, a large-scale collection of common sense knowledge[4]. A list of emotion words (named ground words) was chosen by which to bind a first set of affective sentences in Open-Mind. These sentences contain other words on which the affective information of the ground

words is propagated, with an attenuation factor. By these new words, a new set of affective sentences in OpenMind is individuated, and so on. This approach can be improved by increasing the number of ground words and by considering the senses of the words. Then, a lexical resource including the relation between affective words and concepts is required.

WordNet-Affect allows us to identify the sentences in OpenMind containing affective words corresponding to affective synsets.

### V.       Elaboration of Romanian and Russian lexical resources

Multiple efforts were made to create lexical resources similar with EnglishWordNet for other languages, e.g., the Italian, Hebrew, Portuguese, and Romanian.[5] Romanian WordNet has been created during BalkaNet [10].

The project aimed at the development of a multilingual lexical database comprising of individual WordNets for the Balkan languages.

Several attempts were taken to develop the Russian WordNet. RussNet is a project of computer thesaurus of Russian vocabulary [1]. An alternative project of Russian version of WordNet is Russian WordNet [2]. Both projects are non-commercial. Two commercial projects aimed to develop wordnets in Russian: RuThes is informational thesaurus used in UIS RUSSIA[6] and Russian WordNet project by the Novosoft company group[7]. Unfortunately, little information is available and even less freely available resources.

We decided to translate WordNet-Affect synsets into Russian and Romanian and, then, create two data sets for learning experiments. The two data sets will be available for research purposes upon publication.

We automatically translated all the words of English WordNet-Affect in Russian and Romanian and manually selected the adequate words that conveyed the same emotion as English ones.

To create the two data sets, we applied three-step approach :

**Automatic translation:** It was done automatically using bilingual dictionaries. On this step, our goal was to

**Table 1: Data sets of affective words: English, Russian, Romanian.**

| Classes | English data | | | | Russian data | | | Romanian data | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | # synsets | % synsets | # words | % words | # words initial | # words | % words | # words initial | # words | % words |
| **anger** | 128 | 21.0 | 318 | 20.7 | 149 | 105 | 13.0 | 316 | 151 | 25.0 |
| **disgust** | 20 | 3.3 | 72 | 4.7 | 46 | 31 | 5.0 | 93 | 43 | 3.9 |
| **fear** | 83 | 13.5 | 208 | 13.5 | 118 | 71 | 14.6 | 123 | 55 | 12.8 |
| **joy** | 228 | 37.2 | 539 | 35.1 | 253 | 183 | 36.2 | 510 | 211 | 37.7 |
| **sadness** | 29 | 4.7 | 309 | 20.1 | 217 | 128 | 25.6 | 241 | 111 | 16.2 |
| **surprise** | 124 | 20.3 | 90 | 5.9 | 54 | 29 | 5.6 | 91 | 48 | 4.4 |
| **Total** | | 100.0 | 1536 | 100.0 | 837 | 547 | 100.0 | 1374 | 619 | 100.0 |

---

4 http://openmind.media.mit.edu

5 http://multiwordnet.itc.it/english/home.php

6 http://www.cir.ru

7 http://research-and-development.novosoft-us.com

obtain as many affective words as possible for analysis. We obtained all the translations for every word in the WordNet-Affect synsets. We decided to exclude all the word combinations, collocations and idioms leaving them for future work. An example of the result obtained for each synset after this step is presented on figure 1.

```
01943022      a:
awed =
      speriat
awestruck =
      cuprins de veneraţie
      cuprins de teamă
awestricken =
      înspăimântat
```

Figure 1. All the translations for every word in the synset.

As it is seen from the example, in the Romanian translation we also obtained word combinations which were extracted from the dictionary.

**Selecting the right translations:** To form the word sets, we checked list of all translations and removed words which meaning was not related to the emotion, e.g., for synset with the meaning "preference" for the word "taste" we removed all translations related to food and left only the word "preferinţa". The example is presented in figure 2.

As we translated every word separately, we obtained a lot of duplicates which have to be removed too. For this step the only restriction was to take the translations that are connected with the emotion of the synset. We postponed part-of-speech correspondence till the later phases.

**Selecting the translation of the synset meaning:** All the words in synset represent one concept, one meaning. The aim of the third step was to find the adequate translation of exactly this meaning. At this step we checked the part of speech correspondence as well.

```
05573914      n:
 preference =
            preferinţă
 penchant =
            inclinaţie
            slabiciune
 predilection =
            predilecţie
 taste =
            a avea gust
            a gusta, a cunoaşte
    a gusta; a degusta (un
    aliment)
            degustare
            fărâmă, bucăţică,
    îmbucătură (de)
            gust
            înclinaţie, prefer-
    inţă
```

Figure 2. The example of translation.

For the Romanian data set, to obtain all synonyms of the translated words, we used the on-line dictionary Dexonline.[8]

This step was the most laborious and difficult. Many

English synsets have quite similar meaning with some nuances. In some cases synsets contained obsolete words which had not find in the dictionary. As it was said above we tended to avoid word combinations, collocations and idioms. However in some cases the exact sense of English synset could be represented only by some combination of Romanian or Russian words. In some cases even English synset was presented word combination. For example:

```
      n#05591681 stage_fright
```

Another example contains German word:

```
n#05600844 world-weariness Weltschmerz
```

In such cases we did not obtain the proper translation. Again, in some cases several English synsets have got the same Romanian or Russian words as translations because we could not reflect the nuances in source language senses in the target languages.

Table 1 contains information about the word number for each of the six WordNet Affect emotions. The columns are designed as follows:

- column # synsets presents the initial number of English synsets with affective labels;
- column % synsets shows its percentage;
- column # words shows the unique words count for the English set;
- column % words shows per cent of words for each emotion;
- column # words initial presents the number of the Russian words after the second step before the selection of the right translation for the each synset;
- columns # words and % words list counts and per cent of the Russian words after the third step;
- columns # words initial, # words, % words list the similar information for the Romanian data set.

## VI.   Conclusion and Future Work

This paper describes the process of Russian and Romanian WordNet-Affect creation. WordNet-Affect is a lexical resource created on the base of Princeton WordNet which contains information about emotions the words convey. It is organized in six basic emotions: anger, disgust, fear, joy, sadness, surprise. WordNet-Affect is a small lexical resource but valuable for its affective annotation. We translated WordNet Affect synsets into Russian and Romanian and, then, create two data sets for learning experiments. The two data sets will be available for research purposes upon publication.

This work is only a first step towards the development of this resource. One of the goals is extending the number of affective words, making use of the predefined WORDNET relations. Another possible way is finding correlations between affective words and words which do not express emotions directly but evoke a clear affect, for example "war" or "murder". This task needs another approach.

## VII.   **VIII. References**

[1]   I. Azarova, O. Mitrofanova, A. Sinopalnik-

---

[8] http://dexoline.ro

ova, M. Yavorskaya, and I. Oparin. Russnet: Building a lexical database for the russian language. In Proceedings of the Workshop on Wordnet Structures and Standardisation and How this affect Wordnet Applications and Evaluation, 2002.

[2] V. Balkova, A. Sukhonogov, and S. Yablonsky. Russian wordnet. In Proceedings of the Second Global Wordnet Conference, 2004.

[3] D. Crystal. Language and The Internet. Cambridge University Press, 2001.

[4] P. Edmonds. Introduction to senseval. ELRA Newsletters, 7(3):337–344, 2002.

[5] P. Ekman. An argument for basic emotions. Cognition and Emotion, 6:169–200, 1992.

[6] Liu, Lieberman, et al. A model of textual affect sensing using real-world knowledge. Proceedings of the 8th international conference on Intelligent user interfaces, Miami, Florida, USA, pages 125 – 132, 2003.

[7] C. Strapparava and R. Mihalcea. Semeval-2007 task 14: Affective text. In Proceedings of the 2008 ACM symposium on Applied computing, 2008.

[8] C. Strapparava and A. Valitutti. Wordnet-affect: an affective extension of wordnet. In Proceedings of the 4th International Conference on Language Resources and Evaluation, pages 1083–1086, 2004.

[9] C. Strapparava, A. Valitutti, and O. Stock. The affective weight of the lexicon. In Proceedings of the 5th International Conference on Language Resources and Evaluation, pages 474–481, 2006.

[10] D. Tufis, B. Mititelu, L. Bozianu, and C. Mihaila. Romanian wordnet: New developments and applications. In Proceedings of the 3rd Conference of the Global WordNet Association, pages 337–344, 2006.