

Concept Lattice Implementation in Semantic Structuring of Adjectives

Potemkin S.

Philological faculty, Moscow State University, Russia
potemkin@philol.msu.ru

Abstract. Methods of the formal concepts analysis (FCA) in application to construction of ontological relations in a class of Russian adjectives characterizing appearance of a person with use of WordNet are discussed. Analysis of their semantic paradigm on the basis of the formal context constructed with application of the bilingual dictionary is made.

Keywords. adjectives, concept lattice, hierarchy, dictionary, human appearance

1 Introduction

In the recent years creation of the computer thesaurus of Russian similar in structure and functionality to WordNet thesaurus [16] attracts large interest [1, 7, 8]. Such thesauri give ample opportunities for investigating semantic relations between the meanings of the words of some Natural Language. Unfortunately, the lexical covering by such thesauri for the languages other than English is limited, despite considerable efforts on sinset expansion and their interrelations (sinset is the basic semantic unit of WordNet; a set of English words which code some semantic value). So a necessity of the automated revealing of lexical-semantic relations from the existing sources, such as test corpora or explanatory dictionaries exists. For the decision of this problem methods of the formal concept analysis (FCA) [11, 13, 14] are involved.

We develop methods using bilingual (English-Russian) dictionaries as a source of the formal context and the further construction of a conceptual network for representation of ontological relations in the class of Russian adjectives.

2 Lexical sources

At revealing the structure of semantic paradigm of certain group of words it is necessary to lean against as full lexical sources as possible. We use:

- The common and special English-Russian dictionaries – the lexical database (LDB) [5].

LDB contains English-Russian equivalents from more than 30 common and special dictionaries, including The English-Russian dictionary (ed. Apresian), Muller's dictionary, electronic dictionaries Lingvo, Poliglossum, Prompt, and many others. Translation dictionaries are exposed to some kind of natural selection as they are daily used by translators for practical purposes, and the bad dictionary are rejected;

- Assessment of a person appearance (dictionary) [2];
- WordNet [16];
- Explanatory dictionaries by Ojegov, Evgenieva, Sharov's frequency dictionary [9];

In this paper we describe the semantic paradigm of the adjectives characterizing appearance of the person. The frequency of the words in this group is rather considerable: *большой* (*big*) - 1631 ipm (items per million), *хороший* (*good*) - 854 ipm, *старый* (*old*) - 528 ipm, *белый* (*white*) - 493 ipm, [9] etc. This group is chosen also in view of its importance for specification of system relations of the Russian rating lexicon, notions about types of lexical values, features of connotation, standard lexical associations [3], understanding the structure of a fiction novel [6]. It is important for lingvo-didactics, as a basis for creation of various manuals for speech developing, training in Russian for the Russians and the foreigners, and also for translation of legal, psychological, etc. documents.

Investigation of the meanings of adjectives is similar to investigation of other parts of speech. The component analysis of adjectives with attraction of explanatory dictionaries is used; corpora research is used for the compatibility analysis of syntagma of type adjective - noun which allows to cluster adjectives as the attributes of certain noun for which some classification [12] is already constructed. Methods of direct in-field testing for revealing connotations, i.e. narrowing the set of possible syntagmatic partners (adjectives) of the given lexeme (noun) [4] are used. System relations in lexicon are reflected in thesaurus where the lexical meaning of an adjective is frequently the same as this of a semantically similar verb or noun.

It seems promising to use bilingual dictionaries and the existing thesauri like Roget's or the widely used WordNet for revealing of semantics of adjectives. The synonymic and antonymic relations between adjectives are developed well enough, however in this area also attraction of bilingual dictionaries essentially enriches lists of synonyms and especially - antonyms [5]. Other types of relations: hyponymy, meronymy, metonymy and so forth are much less investigated. Revealing of the specified relations between adjectives is of theoretical and practical interest, especially in application to the Automatic Text Processing and Natural Language Understanding. In this case the direct support on the WordNet structure is unproductive. Really, that the semantic organization of qualitative adjectives in WordNet completely differs from the semantic organization of nouns or verbs. Adjectives are organized in clusters linked to a "focal" adjective having an antonym, i.e. antonymic relation is the base semantic relation for coding meaning of adjectives. This approach is connected with the fact, that adjectives have attributive function and that a considerable number of attributes are bipolar. No hierarchical relations similar to the hyponymy relations between nouns or troponymy relations between verbs are revealed in WordNet for adjectives and, as a rule, the direct hypernym is not indicated, instead of it the refer-

of it the reference «Pertains to noun ...» is given, that hypernym of an adjective often is a noun, for example for the adjectives designating size (*big, small, narrow, spacious*) a generic hypernym is the noun "size". In this paper we expect, however, to find hierarchical, etc. relations within the class of adjectives.

3 Formal Concept Analysis (FCA)

The formal concept analysis is based on intuitive guess that concept has two parties: *an extent* which contains some objects, and *intent* which includes all attributes peculiar to these objects [16]. For the formal analysis of concepts it is necessary to define, first of all, *a formal context*, $K: = (G, M, I)$, where G = set of objects; M = set of attributes; and I = the binary relation between elements of G and M , showing, what attributes m are attributed to objects g . It is easy to present a formal context in the form of a table. Table 1 contains some adjectives of Russian as objects, a set of translations of these adjectives – as attributes; the certain Russian word, e.g. *алчный* has a translation equivalent *rapacious*, crossing of the corresponding line and column is marked by cross (X). Derivation operation over the formal context is defined as follows:

$$X \subseteq G: X \rightarrow X': \{m \in M | gIm \text{ for all } g \in X\}; Y \subseteq M: Y \rightarrow Y': \{g \in G | gIm \text{ for all } m \in Y\}$$

In our example let $X: = \{\text{ХИЩНЫЙ}, \text{прожорливый}\}$ and let $Y: = \{\text{ravening, wolfish}\}$. Then $X' = \{\text{ravening, rapacious, ravenous}\}$, $Y' = \{\text{ХИЩНЫЙ}, \text{жадный}\}$, further $X'' = \{\text{ХИЩНЫЙ}, \text{жадный, прожорливый}\}$, etc. It is possible to show that generally $X \subseteq X''$ and $X' = X'''$ and also $Y \subseteq Y''$ and $Y' = Y'''$. The *formal concept* for the given formal context is the pair (A, B) where $A = B'$, $B = A'$, i.e. A = set of objects, having all attributes from the set B , B = set of attributes attributed to all objects of the set A . All formal concepts for the given formal context are generated as (X'', X') or (Y', Y'') , for all subsets $X \subseteq G$ or $Y \subseteq M$. A number of algorithms for the fast construction of formal concepts are developed [15]. The cells representing formal concept (A, B) are highlighted in our table; $A = \{\text{алчный, грабительский}\}$; $B = \{\text{rapacious, ravenous}\}$. Relation \leq establishes a partial order over the formal concepts for the given formal context $B(K): (A_1, B_1) \leq (A_2, B_2)$. $\leftrightarrow A_1 \subseteq A_2 (B_2 \subseteq B_1)$. This relation is called as the relation *subconcept – superconcept* and \leq defines a complete lattice $\underline{B}(K)$ over $B(K)$ which can be depicted in the form of the labeled oriented graph (fig. 1). The nodes this graph are the formal concepts, and the edges reflect the *subconcept – superconcept* relation.

We propose to use thesaurus WordNet and FCA methods to reveal semantic paradigm of Russian adjectives. Basic semantic unit of WordNet is a synset - a set of English words which in aggregate code some semantic meaning. An element of synset is word

meaning (WM) - the meaning of a single word (word-combination), included in a synset.

Table 1. the Formal context for a synset. The objects from the Dictionary are capitalized.

	edacious	esurient	ravening	rapacious	ravenous	voracious	wolfish
ЗВЕРИНЫЙ							X
ЗВЕРСКИЙ							X
СВИРЕПЫЙ							X
ХИЩНЫЙ			X	X	X		X
Алчный				X	X		
Грабительский				X	X		
Волчий					X		X
Голодный		X			X		
голодный как волк					X		
жадный	X	X	X	X	X	X	X
жаждущий					X		
захватнический				X			
изголодавшийся					X		
ненасытный		X		X	X	X	
относящийся к волкам							X
очень голодный					X		
падкий						X	
похожий на волка							X
прожорливый	X	X	X	X	X	X	
свинский				X			
характерный для волка							X
эгоистичный				X			

A word can participate in various synsets, that reflects polysemanticism and homonymy (homography) inherent in the given word. Synsets participate in hypo – hypernymic relations (for nouns), troponymic relations (for verbs), antonymic, meronymic relations and so forth. Synsets, containing adjectives, as a rule, are not captured by hyponymy relations, establishment of hierarchical relations between adjectives is hard both from the theoretical and practical points of view [1,12]. Nevertheless, using synsets for revealing of semantic paradigm of adjectives is obviously possible and promising. We note, first of all, that the bilingual English-Russian dictionary can effectively be applied to expansion of the list of synonyms, and also definition of semantic affinity among Russian synonyms [5]. It is possible to assume, that taking a set of the English words of a synset, $\{e_i\}$, i.e. synonyms with certain meaning, and all their translations into Russian $L_j(e_i) = r_{ij}$, intersection $\bigcap_{ij} r_{ij}$ will contain a set of the Russian words coding meaning, equivalent to the synset $\{e_i\}$ meaning. Owing to various reality partitioning in English and Russian which is the direct reflection of discrepancy of the category assignment and, hence, concept assignment of attributives,

and also propensity of English to the greater detailing of the world a nomination of various features, such intersection as a rule, is empty, or contains several words with very wide semantics. Therefore we propose to use FCA which will allow revealing the whole structure of sets $\{r_i\}_j$ in their interrelation with synset $\{e_i\}$. Formal context $K: = (G, M, I)$ in this case consists of a set of objects $G = \cup_j \{r_i\}_j$ of all translations of all English words from a synset; set of attributes $M = \{e_i\}$; the binary relation I is defined by attaching the Russian equivalent j to each English word e_i (Table 1).

4 Experimental results and interpretation

The experimental approbation of our technique was carried out over the Dictionary « Assessment of a person appearance» [2], (hereinafter - the Dictionary) containing more than 200 dominants and more than 1200 members of synonymic series of the adjectives attributed to appearance of a person. In particular, 603 adjectives for which more low 1040 conceptual lattices with number of attributes more than 2 have been constructed. For each adjective ar_i all English equivalents $ae_{ij}=L_j(ar_i)$ from the Dictionary containing in the lexical database (LDB) are listed. For every ae_{ij} the set of synsets $\{s_k\} = WN(ae_{ij})$ containing ae_{ij} is defined. For each synset s_k all Russian adjectives which are the translation equivalents of the synset elements are listed; doubles are rejected. Thus, the set of objects G and a set of attributes M of formal context K are received. At this stage we do not carry out the semantic division of inconsistent translation equivalents (which actually exist, e.g. *large-handed* it is translated as *жадный* and as *расточительный*). Also the adjectives concerning appearance of the person are not selected; such selection is carried out later, at an analysis stage of the constructed conceptual lattice. All pairs of equivalents are included in the Table.

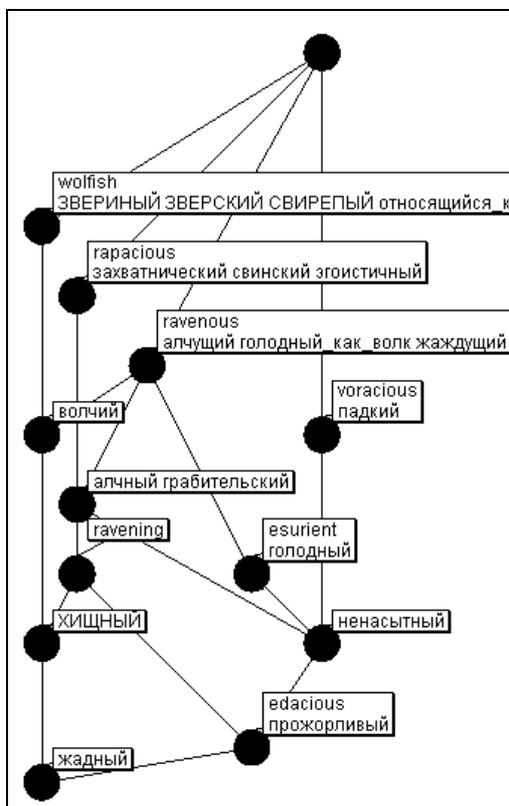


Fig. 1. Conceptual lattice over a formal context of Table 1.

Within the framework of synset №00011320 object ХИЩНЫЙ is a hypernym for objects ЗВЕРИНЫЙ, ЗВЕРСКИЙ, СВИРЕПЫЙ. Such definition of a hypernym generally is not seems to be correct (*зверь (animal) is not necessary хищник (predator)*, see Efremova []: *зверь1 = Wild, usually predatory animal*), but as the characteristic of the person the beastly, brutal, furious person most likely is the predatory person. The following hyponymy relations are revealed while analyzing other synsets:

- мертвый (dead) ⊆ неподвижный (motionless) ⊆ вялый (languid)*
- апатичный (apathetic), оцепенелый (frozen) ⊆ вялый (languid)*
- изящный (graceful) ⊆ тонкий (delicate)*
- коварный (artful) ⊆ хитрый (sly)*
- нахальный (impudent), самоуверенный (self-confident) ⊆ дерзкий (daring) ⊆ смелый (brave)*
- решительный (decisive) ⊆ твердый (hard)*
- ястребиный (hawk) ⊆ хищный (predatory)*
- мерзкий (vile), отвратительный (disgusting), противный (offensive), ужасный (awful) ⊆ неприятный (unpleasant)*

Some of these relations coincide with those registered in the Dictionary: *изящный (graceful) ⊆ тонкий (delicate)*, *коварный (artful) ⊆ хитрый (sly)*, the others are

newly revealed, or contradict the Dictionary, e.g. in the Dictionary adjective *ястребиный* (*hawk*) is a hyponym of the adjective *белчий* (*squirrel*) (?).

Using FCA it is also possible to find adjectives attributed to the human face which could enter the Dictionary: *бесчувственный* (*insensible*), *будничный* (*every day*), *выцветший* (*faded*), *загадочный* (*mysterious*), *заспанный* (*sleepy*), *зловещий* (*ominous*), *искаженный* (*deformed*), *легкомысленный* (*thoughtless*), *матовый* (*matte*), *незамысловатый* (*plain*), *нездоровый* (*unhealthy*), *неприметный* (*imperceptible*), *плоский* (*flat*), *полусонный* (*dozing*), *придурковатый* (*foolish*), *притворный* (*feigned*), *разбойничий* (*predatory*), *смущенный* (*confused*), *сухощавый* (*lean*), *флегматичный* (*phlegmatic*), *худой* (*thin*)...

Also the attributive word-combinations which are not included in the Dictionary at all are revealed: *с буйной растительностью* (*with the violent vegetation*), *наводящий скуку* (*boring*), *с хитрецей* (*sly*) ... Comparison of all received hierarchical relations to the Dictionary is out of scope of this research. The proposed method has only allowed to reveal additional lexical units and to establish semantic relations which can be used both in lexicography, and for Automatic Text Processing.

5 Conclusions and research prospects

Complexity of the problem of revealing semantic structure of adjectives is confirmed by the previous researches. Application of methods of the formal concept analysis (FCA) for its decision can appear useful as addition to the corpora – based methods, the component analysis, etc. It is supposed to develop the described methods for formal revealing hierarchical relations from the concept lattice. Besides, expansion of the proposed approach on other semantic relations is possible.

References

1. Azarova, I.V., Sinopalnikova, A.A., Javorsky, M.V. Principles of construction of WordNet-thesaurus RussNet (in Russian) In: *Computer linguistics and intellectual technologies. - Proceedings of the International conference Dialogue'2004* pp.542–547. Moscow (2004)
2. Boguslavsky, V.M. Assessment of appearance of a person, Dictionary. (in Russian) Publishing house "Ast" Moscow (2004)
3. Kedrova, G. E, Potemkin, S.B. Semantic discrimination of homonyms using bilingual dictionary and dictionary of synonyms (in Russian) In: *Proceedings of II International congress "Russian: historical destiny and the present"*, Moscow. (2004)
4. Kobozeva I.M. (2000) Linguistic semantics publishing house «Editorial UPCC», M. 2000, 350 pp.
5. Potemkin, S.B. Lexical database with the imposed semantic metrics (in Russian). In: *Proceedings of II International congress "Russian: historical destiny and the present"*, Moscow (2004)

6. Potemkin, S.B. Detection of event by analysis of antonyms in N.V.Gogol and A.P.Chehov's texts. (in Russian) In: *The word and the dictionary - Proceedings of the International scientific conference «Modern problems of lexicography»*, pp.93-95, Grodno (2009)
7. <http://www.cir.ru/>.
8. Sukhonogov, A.M. Yablonsky, S.A. (2004) Automation of English-Russian WordNet construction. (in Russian) In: *Proceedings RDCL 2004. September, 29 - October, 1*. Pushino (2004).
9. <http://www.artint.ru/projects/frqlist.asp>
10. Javorsky, M. B, Azarov, I.V. Structure of attributive meanings in RussNet thesaurus. (in Russian) In: *Proceedings of the International conference Dialogue'2009* pp.542-547 Beka-sovo (2009)
11. Cimiano, P, Hotho, A., Staab, S. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. In: *Journal of Artificial Intelligence Research*. Volume 24, p.305-339 (2005)
12. Mendes, Sara Adjectives in WordNet. In: *PT//GWC 2006, Proceedings*, pp. 225-230. (2006)
13. Priss, U. Linguistic Applications of Formal Concept Analysis. In Ganter; Stumme; Wille (eds.), *Formal Concept Analysis, Foundations and Applications*. Springer Verlag. LNAI 3626, pp. 149-160. (2005)
14. Stepanova, N.A. Automatic acquisition of lexical-semantic knowledge from corpora. In: *SENSE'09 Proceeding shop* pp.91-100, Moscow (2009)
15. Wille, R. Restructuring lattice theory: an approach based on hierarchies of concepts. In: Rival, I. (ed.) *Ordered Sets*. p.445-470. Dordrecht-Boston, (1982)
16. Fellbaum, Ch. (ed.) WordNet: An Electronic Lexical Database. MIT Press. (1998)