Hierarchy of perceptional adjectives in RussNet

Irina Azarova
St-Petersburg State University /
Universitetskaya nab. 11, 199034
St-Petersburg, Russia
ivazarova@gmail.com

Maria Yavorskaya
St-Petersburg State University /
Universitetskaya nab. 11, 199034
St-Petersburg, Russia
Yav.mas@gmail.com

Abstract

The group of perceptional adjectives in RussNet is presented. The structuring of adjectival synsets follows different patterns: bipolar clustering, hyponymy, and troponymy. Several complex problems of hyponymy linking are discussed.

1 Introduction

Adjectives is a class of ambivalent words, which may share features with nouns or verbs forming noun-verb continuum (Wetzer, 1996), that is, some are closer to nouns, others — to verbs. The difference is expressed in grammatical behaviour: attributive-only, predicative-only or attributive-and-predicative (Quirk et al, 1985) syntactic usage and derivational productivity — forming degrees of quality or intensifications/ diminutives, - that is considered in Russian grammatical tradition to be the evidence of their descriptive origin.

The semantic approach to adjectives is based on prototypical semantic classes (Dimension, Age, Value and Colour) (Dixon, 2004) and other miner classes denoting difficulty, similarity, quantification, position, etc. The attribute expressed by an adjective refers to a wide or narrow class of objects, so it may more or less definitely combine the concept of quality with that of an object features. For example, Russian adjective *rusyj* (*fair*) clearly specifies the object class — hair (or a human with such hair), and its definition 'light brown' in the explanatory dictionary looks incorrect (the gloss for *fair* in WordNet is more adequate: 'pale yellowish to yellowish brown'). This quality of descriptive adjectives affords them to substitute a noun phrase, and they are considered to be conceived on a class of objects as its domain.

A sequence of adjectives may specify the intersection of classes as in *big brown shoes* (intersection of a class of big shoes with that of brown shoes). In relation to logical behaviour of adjectives, they are differentiated as intensional and extensional (intersective vs subsective) (Peters & Peters, 2000). The construction of a noun with several adjectives is normally represented by a positional pattern as in an example above, confer *brown big shoes*.

The usual structuring of descriptive adjectives word meanings is considered to be "bipolar clusters" (Miller, 1998), that is, pairs of focal adjectives form the centres of opposite poles for attribute values, a number of descriptive adjectives being associated to centres by semantic similarity. This pattern is very typical for different wordnets, however, in (Azarova and Sinopalnikova, 2004) it was shown that in some groups of adjectival meanings the proper hierarchy may be set up. In some cases it looks like noun hyponymy, e.g. *bolshoj1 (big)* 'having great or greater than average spatial extension' has hyponyms *dlinnyj1 (long)* 'having great or greater spatial extension between two extreme points in *horizontal* direction' and *vysokij2 (high)* — 'having great or greater spatial extension between two extreme points in *vertical* direction', etc. The meaning of a hypernym is specified in a particular manner (an exact direction) in hyponyms.

In this paper we deal with adjectives from the semantic field of Perception. We suppose that this group represents the core elements of our cognitive space and a hypothetical hierarchy may be outlined in this domain. The most frequent Russian adjectives from (Sharoff, 2001) with reference to different spheres of perception: visual, auditory, tactile, gustatory, and olfactory.

The RussNet technique (Azarova et al., 2005) of data processing involves marking up the

random samples of 150-300 contexts from the balanced corpus of contemporary texts in order to extrapolate the frequency distribution of word meanings to the general aggregate of adjective occurrences in the corpus. This procedure allows us to order word meanings according their frequencies in the format *bolshoj1, bolshoj2, etc.* (in contrast with traditional dictionary enumeration of meanings, which is rather subjective and sometimes arbitrary). We'll use an abbreviation WM for this specified pair word-meaning, and measure a WM frequency of co-occurrences in *ipm* (instances per million words in the corpus). Experiments with distributions of word meanings in the corpus showed that WMs with ipm-index more than 60 are concepts from the top ontology, and less than 0.1 are potential, because a list of these rare WMs tends to be unlimited. The ipm-index is very helpful as an evidence in hypernym-hyponym relation. The sum of all processed WMs in the adjectival perception group is about 3,500 ipm, that covers 73,500 corpus contexts.

We use also the following dictionaries: the Dictionary of Russian in 4 volumes (Evgenjeva, 1981–1984) for preliminary word meaning definitions, the Russian Associative Dictionary (Karaulov et al., 2002) for checking semantic links by means of an associative rank (the number of similar responses to the word stimulus).

We suppose that WMs of frequent adjectives outline the general structure of perceptional domains. In order to prove its comprehensibility we process a number of rare WMs. We'll demonstrate that there are several modes of group structuring, the special attention being focused to the types of hypernym-hyponym relation.

2 Hyponymy link

The hyponymy linking between a pair of WMs is based on 3 evidences.

The first one is dictionary definition, for example, *rozovyj1* (rose-coloured, pink) 'pale red' (Evgenjeva, 1981–1984), the colour is defined as a tint of red, that is, a variant or a hyponym.

The second is the corpus evidence: the hypernym and the hyponym are to have similar collocations, though the hypernym has wider usage, and, consequently, higher ipm-index. *Krasnyj1* (red) ipm is 120.59, and *rozovyj1* (rose-coloured, pink) — 17.31, which conforms to requirements of ipm-indices correlation between the hypernym and hyponym. The value of ipm for *krasnyj1* indicates that this concept is from top ontology (as well as *belyj1* 'white' – 152.44, *chernyj1* 'black' – 138, and *zelenyj1* 'green' – 82.7).

The third evidence is used in doubtful cases, it is the association rank in (Karaulov et al., 2002), usually there is a response with highest rank or one the upper part of the article. *Rozovyj1* (rose-coloured, pink) has a response *krasnyj1* (red) with a rank 3. Association response is usually oriented from a hyponym to the hypernym. This association rank is not the highest, which is 17 for a noun *tsvet* (colour).

The 3 sources may disagree: the first and the second are in favour of hyponymy linking of *rozovyj1* and *krasnyj1*, but the third one hardly supports this relation.

It may be easily seen that the hyponymy in this pair is of a troponymy type (as in a verb pairs *walk–strike*), it is peripheral realisation of an action or a quality (in the case of adjective WMs).

The different instance is a relation between *bol'shoj1* (big) and *ogromnyj1* (huge) 'very big', the latter being the realisation of a lexical function MAGN (Zholkovskij & Mel'chuk, 1965) — intensification of an attribute, confer other functions: ANTIMAGN, BON and ANTIBON. In this pair all evidences are in favour of hyponymy linking. An association response for *ogromnyj* is *bol'shoj* with rank 40. The realisation of a lexical function MAGN is stylistically marked, this link is typical for attachments of expressive designations in bipolar clusters, which is a traditional wordnet relation in attributive structuring. The difference is in direction of relations: similarity presupposes symmetry, which is not proved by RussNet data.

The third type of hyponymy in attributive structures is logically similar to true hyponymy. For example, *vysokij2* (high) and *vysokij6* (tall), the latter in relation to the former is subsective in a SIMPLE sense (Peters & Peters, 2000), that is, *vysokij6* refers to the subclass of that for *vysokij2*.

2.1 The controversial points of hyponymy

Normally the hyponymy relation presupposes that synsets are from one POS, the meaning of a hyponym incorporates the meaning of the hypernym and its specification.

The first disputable point of adjectival hyponymy is that the top-level concept for some parameter specification may be represented by a noun WM. This is manifested by the highest

association rank of a noun WM *tsvet* (colour) in an example above.

We can use the unilateral implication test for verification of the semantic relation: let B be a hyponym for A, then

A={*vkus (taste)*}, B={*kislyj2 (acid)*} =>
C=*The fruit is* _____.

T*he fruit is acid* implies that *the fruit has (certain) taste* (true). But *the fruit has taste* doesn't imply that *the fruit is acid* (false).

The second point concerns the uniqueness of the hypernym. For example, the adjective WM *seryj1* (grey) according to (Evgenjeva, 1981–1984) denotes 'of a colour resulted from blending black and white, as of ashes', thus WM *seryj1* (grey) has 2 different hypernyms *chernyj1* (black) and *belyj1* (white). This is an intersection in a monochromic designation, but similar connections are very common in a structure of coloured attributes. For example, *ognennyj2* (flame-coloured) is defined as "orange-red", where the main colour is red (it is a standard superordinate notion of the troponymy type described above), but the second one (*orange*) is a troponym of *red.* So two hypernyms (or co-hypernyms) may be from the same level of hierarchy (intersective) or from hypernym-hyponym levels of hierarchy
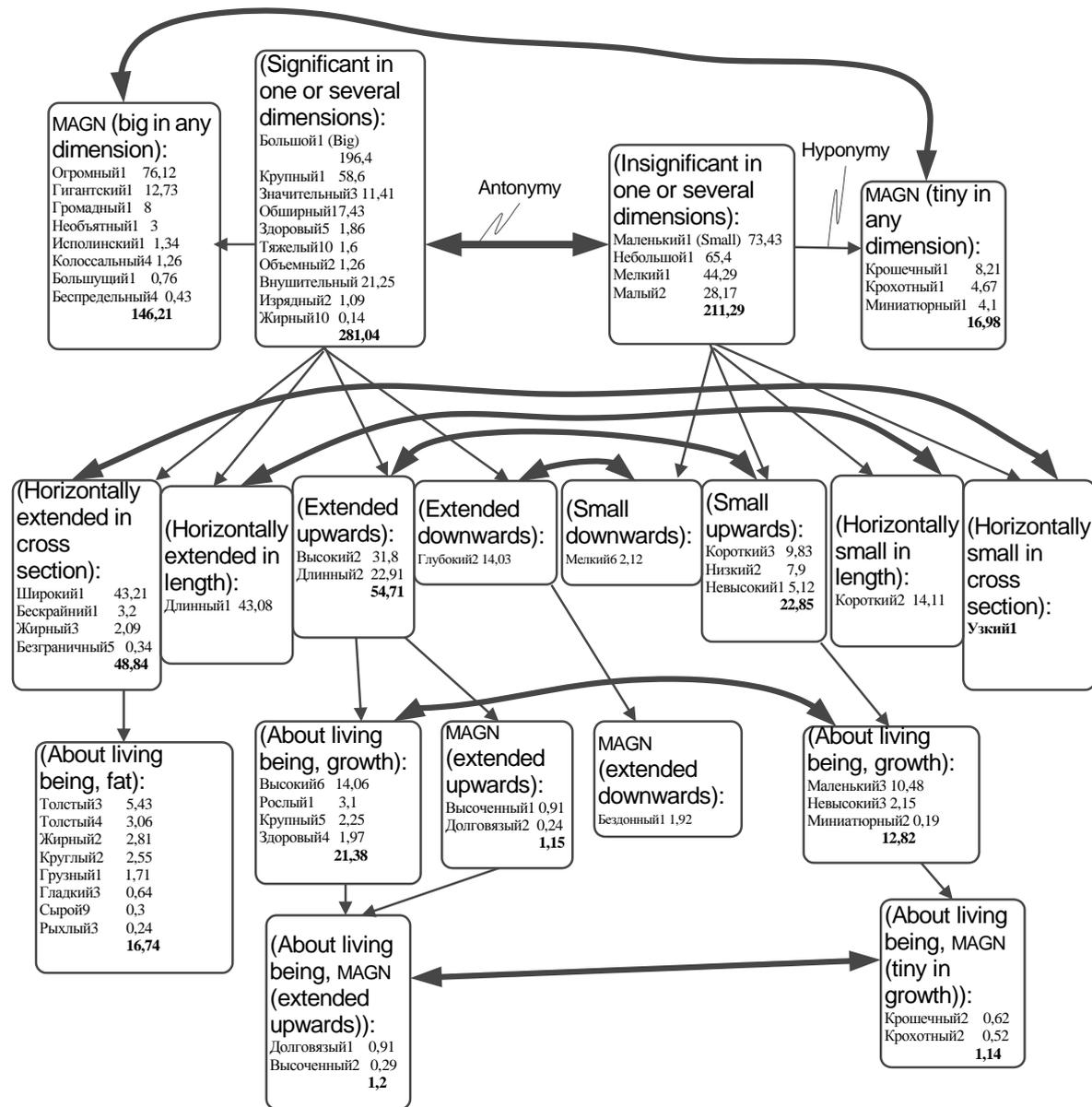

Figure 1. Adjectives of size.

(subsective). We propose to make weaker the connection with the subordinate concept *orange* inserting a prefix into the relation type — LIGHT-HYPONYMY.

The third point is connected with a prototypical pattern for an attribute, as is the case with a description of *seryj1* (grey) '…as of ashes'. It may be useful for text interpretation procedures to gain information of this type. In a way this link is similar to INVOLVED or ROLE relations.

Thus, this *seryj1* (grey) is connected to noun WM *pepel* (ash) with a link ATTRIBUTE.

3 Adjectives of visual perception

This group includes such parameters as size, colour, shape, surface property etc., the characteristics perceived by eyesight. It is the most elaborate structure of different perceptional spheres.

3.1 Adjectives of size

Size is a prototypical concept field for attributive meanings. Fig.1 represents semantic relations in this group. Squares configure synsets, its members — WMs — are ordered according to their ipm indices. The number of line shows the place of a WM in the synset, the bold total at the bottom of a square shows the ipm index of a synset. Arrows indicate semantic relations between two synsets. One-side links mark the link from the hypernym to hyponyms, two-sides links are used for symmetrical relations — antonyms.

The top synset *bolshoj1 (big)* has antonymic synset *malenkij1 (small)* and is adjoined by several types of hyponyms.

MAGN synsets *ogromnyj1* (*very big, huge*) and *kroshechnyj1 (very small, tiny)* are linked with *bol'shoj1 (big)* and *malen'kij1 (small)* as traditional adjectival clusters.

Another type of hyponymy is represented by 8 synsets (*shirokij1* (*wide*) — *dlinnyj1 (long)* — *vysokij2 (high*) — *glubokij2 (deep)* and their antonyms: *uzkij1 (narrow)* – *korotkij2 (short)* – *korotkij3 (low)* – *melkij6 (shallow))* and specifies size according to one of the dimensions, so their co-occurrence is narrower than that of the hypernym (*deep column) anyway all of them are visible objects. The relation between superordinate and subordinate synsets is of the noun hyponymy type, they are subsective.

The third type of hyponyms is represented by very specific co-occurrence with nouns. For example, some synsets have attributes defined only over the class of living beings (*dolgovyazyj1 (lanky)* or *kroshechnyj2 (tiny – about creatures)).* This type resembles both
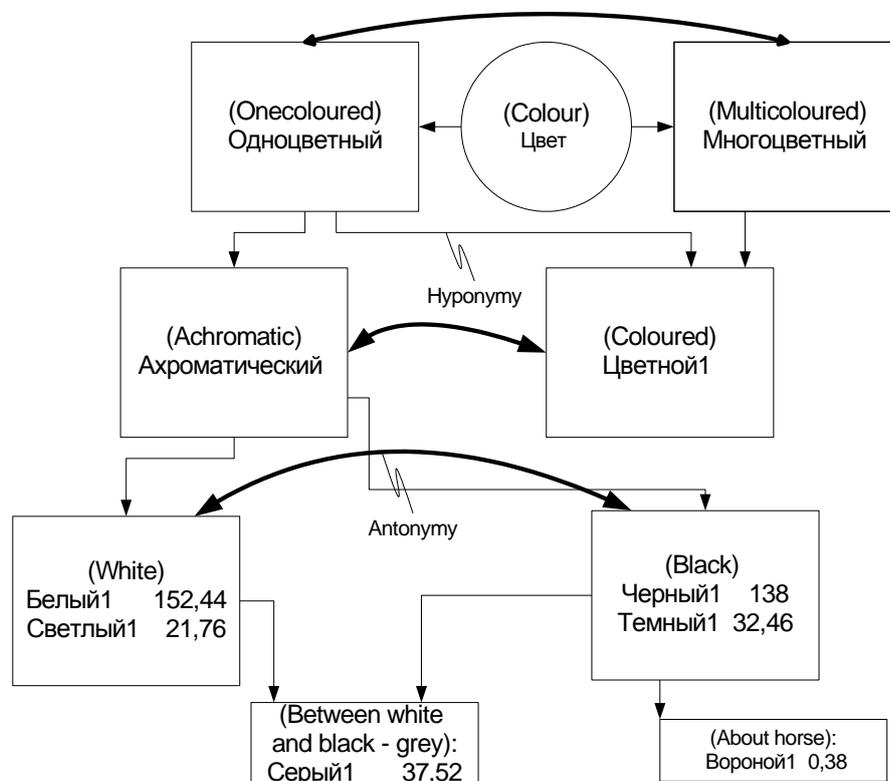
Figure 2. Adjectives of colour. Top of the hierarchy.

hyponymy of nouns and troponymy of verbs.

3.2 Adjectives of colour.

Colour is another prototypical field of adjectival meanings. Fig.2 represents the top levels of the Colour structure. In this structure there are two substructures: black and white structure (achromatic) and coloured attributes (chromatic). The root of the structure is represented by the noun WM (colour). Antonymy exists only in monochromic substructure.

Synsets *achromaticheskij (achromatic)* or *mnogotsvetnyj (multicoloured)* are low frequent and inserted to conform with scientific picture of the world, they may be considered virtual, because there is no true lexicalised expression for them. They are inserted to support the hierarchy, meanwhile they make more convenient representation of mixed colours or merge of colour with a black and white palette: *cherno-belyj, cherno-krasnyj (black-and-white, black-and-red)* etc.

Fig.3 represents a part of the "coloured" hierarchy with brown tints. This structure includes troponymy-like hyponyms (*red + yellow + brown +...=coloured*) and hyponyms defined over specified semantic classes (*hair, horse, meal etc.*)

In this structure there are examples of LIGHT HYPONYMY: a synset belongs to one co-hypernym MORE than to the other. Ex.: *rumyanyj2* is defined as "glittering gold-and-brown, about roasted meal". The main hypernym is obviously *brown* but we can't ignore *glittering gold*, although *rumyanyj2* doesn't belong to the yellow tints. The new type of hyponymy is marked by dashed and dotted arrows on the Fig.3 (LIGHT_HYPONYMY).

In this structure WM for adjectives of colours are connected with noun synsets as a prototypical representatives by means of a link of a new type ATTRIBUTE: *belyj1 (white)* is defined as "of a colour of milk, snow, chalk" or *fistashkovyj2 (pistachio)* has a definition "of a colour of pistachio nut nucleus".

In some projects our colleagues prefer to regard relations between nouns, denoting these objects, and adjectives of colour as regular polysemy (Barque et al. 2008). We can't agree with them. First of all, polysemy is a relation between members of the same POS, so adjectival synset {*cerise1, cherry4*} just can't be polysemic to the noun {*cherry*}. The real reason for this may be the inflectional character of Russian, in which adjectives differ from nouns in form (cf. *vishnevyj3 (cerise) and vishnya (cherry)*). Obviously, such objects are prototypical for the
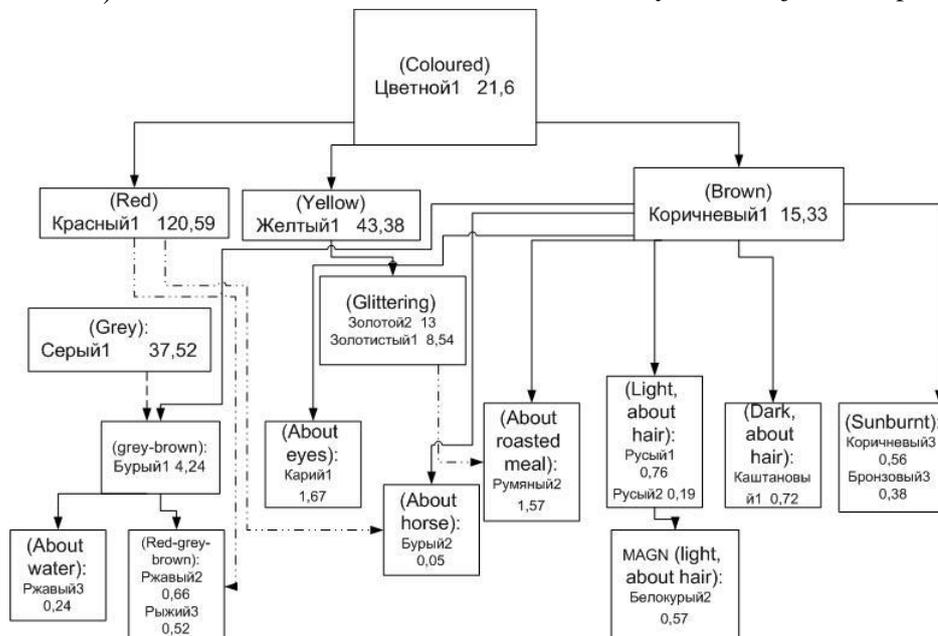

Figure 3. Adjectives of colour. Brown tints.

colours and there has to be a relationship between them. The primary colours are not motivated from the contemporary point of view, and the existed relation XPOS_NEAR_SYNONYMY is not applicable, so ATTRIBUTE link would be useful.

4 Adjectives of auditory perception

It is common knowledge that Sound can be estimated in terms of its volume, pitch (diapason), time-value, purity and site of generation. It's curious that according to frequency the main dichotomy here is not the pair *gromkij1-tikhij1 (loud-quiet)* but *dlinnyj5-korotkij6 (long-short).*

It ought to be noted that the whole group is though numerous but rather low frequent. Even the dominant doesn't exceed 17 ipm. Nevertheless, we plan to structure it in the similar way.

5 Adjectives of olfactory perception

This group of adjectives in this group is concise, and low frequent. Its total ipm index doesn't exceed 30 ipm. This fact can be easily explained by a very productive syntactic phrase in Russian "*smell of N*" which may describe practically every odour without adjectival elements. The specific feature of adjectives in his group is a representation of lexical functions BON and ANTIBON (attractive or not), MAGN and ANTIMAGN (strong or weak). Sometimes components are specified.

Fig.4 represents the structured subgroup of smell attractiveness. There are also such individual synsets as "very strong smell", "combined smell" and "smell that leaves certain sensation in the mouth" (*solonovatyj2 – slightly salty smell*).

6 Adjectives of gustatory perception

Fig.5 represents adjectives of taste, where noun-hypernym is crucial to create any structure at all. Such adjectives as *bitter, sweet, salty etc.* have nothing in common but certain sensation they leave on a tongue that means Taste. Core tastes differ from each other by collocation with noun classes. Sometimes, *sladkij* (sweet) and *kislyj* (sour) are considered to be antonyms, though opposition in contexts is sporadic: in combination with a noun *jabloko* (apple).

Moreover, the data of (Karaulov et al., 2002) show no sign of antonymic opposition in association responses.

But the ultimate argument from our point of view is a composite word *kislo-sladkij* (sour and sweet).

The group is to the great extent alike the adjectives of colour:

- there are several main tastes and their shades, some of which can be described by means of co-hypernymy.

- We can't clearly distinguish tastes and in several cases are forced to rely on dictionary definitions.

- Antonymy is rare and disputable. We should decide by ourselves if any pair like "*light*
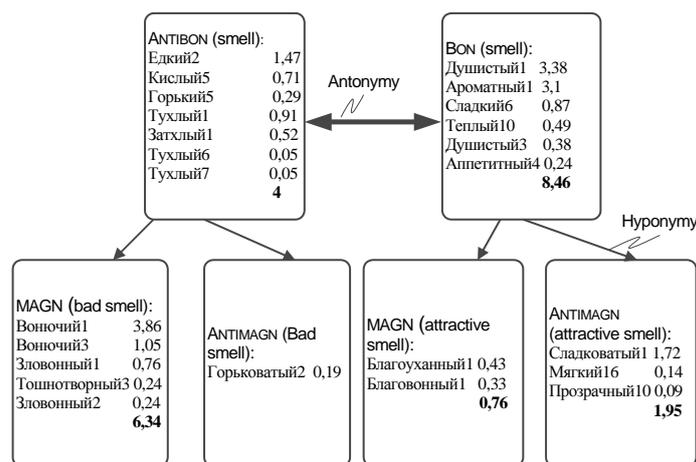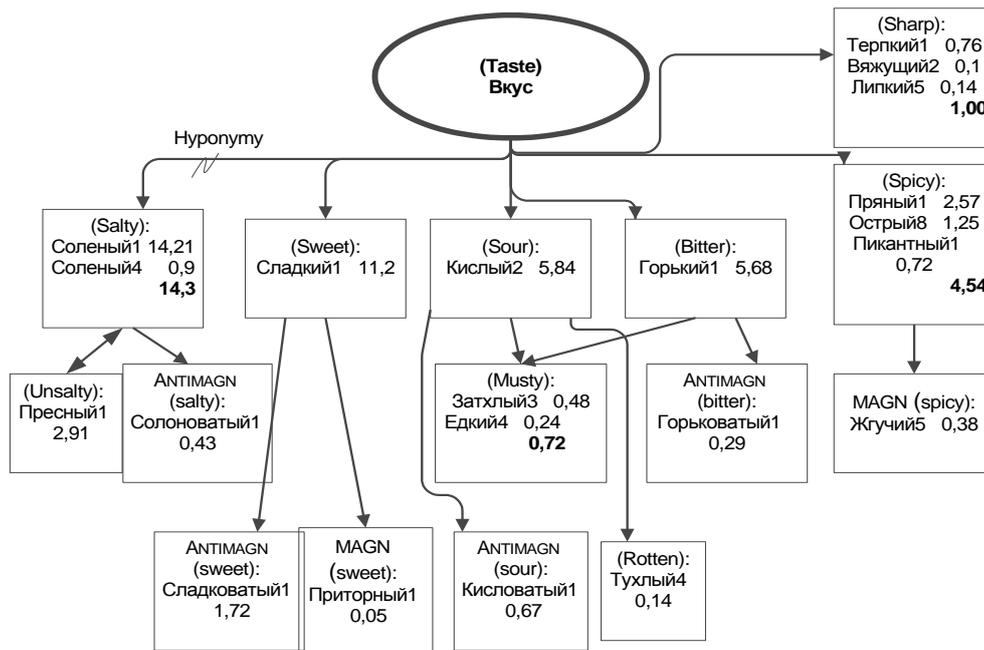

Figure 4. Adjectives of odour. Attractiveness.

Figure 5. Adjectives of taste.

attribute" – "*strong attribute*" demonstrate antonymy or not. And the decision is pragmatic. For example, we don't regard WM *sladkovatyj1* (slightly sweet) and WM *pritornyj1 (*very-very sweet*)* as antonyms, neither we did in other subgroups: *blagouhannyj1* (very attractive smell) – *sladkovatyj1* (slightly attractive sweet smell).

But due to the fact that taste perception is less verbalised and perhaps the human classification of the taste itself is less elaborated than that of visual images, we don't expect to meet light hyponymy in this group.

7 Adjectives of tactile perception

This group is divided into adjectives of weight, heat, moisture, characteristics of the whole body (*myagkij1, lomkij1 – soft, fragile*) and characteristics of its surface (*skol'zkij1, pushistyj1, lipkij1 – slippery, fluffy, sticky*) – everything we can feel by our skin.
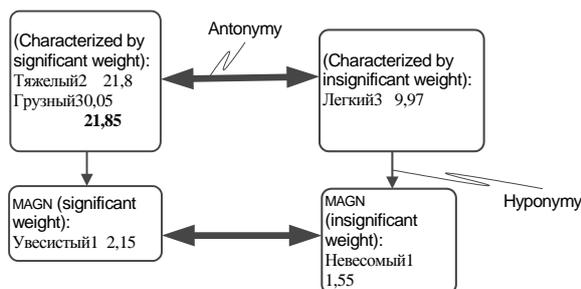

Figure 6. Adjectives of weight.

Some of these subgroups, like adjectives of weight, have rather obvious hierarchy: see Fig.6.

Others, like characteristics of surface, seem to be very flat in structure with variety of troponym-like hyponyms.

8 Conclusions

As we have seen, the groups of perceptional adjectives are not uniform both on the level of their lexicalisation in Russian and on the level of a number of synsets and the depth of hierarchy of subgroup structuring.

Nevertheless, we have shown that in many cases we just can't be content with the conventional point of view, which doesn't foresee any type of the hyponymy in adjectival structures. Moreover, we demonstrated types of hyponymy for adjectives, which combine the hyponymy for typical nouns and the troponymy for verbs with several particularly adjectival features:

- noun-like classical HYPONYMY
- specific mainly for verbs TROPONYMY.
- CO-HYPERNYMY (there exit two hypernyms, for one hyponym)
- LIGHT HYPONYMY (there exit two hypernyms, one of which is the main and the other is the secondary for the hyponym).

We hope that proper structuring of perceptional adjectives will help to represent the

whole group of adjectival WMs in RussNet database.

References

Alonge, Antonietta & others. Encoding information on adjectives in a lexical-semantic net for computational applications. http://www.aclweb.org/anthology-new/A/A00/A00-2006.pdf

Azarova, Irina and Sinopalnikova, Anna (2004) Adjectives in RussNet. *Proc. of the 2nd Global WordNet Conference.* Brno: Masaryk University. Pp. 251–258.

Azarova I.V., Ovchinnikova E.A., Ivanov V., Sinopalnikova A. A.. (2005) RussNet as a Semantic Component of the Text Analyser for Russian. In *Proceedings of the Third International WordNet Conference*. Brno Masaryk University. Pp. 19–28.

Barque, Lucie and Chaumartin, François-Régis. (2008). Regular Polysemy in WordNet.

Dictionary of Russian in 4 volumes (1981–1984) Ed. A.P.Evgenjeva, Moscow.

Dixon R.M.W. (2004) Adjective Classes in Typological Perspective. In R.M.W. Dixon and A.Y. Aikhenvald, Eds. Adjective Classes. A Cross-Linguistics Typology. Oxford: Oxford University Press, pp. 1-49.

Mendes, Sara. (2006). Adjectives in WordNet. *Proc. of the 3nd Global WordNet Conference*: pp.225-230.

Miller, Katherine J. (1998). Modifiers in WordNet. In C/Fellbaum, Ed. *WordNet. An Electronic Lexical Database*. MA, The MIT Press. Pp. 47-67.

Naumann, K. (2000). Adjectives in GermaNet. http://www.sfs.uni-tuebingen.de/GermaNet/Adj.html

Peters, Ivonne and Peters, Wim. (2000). The Treatment of Adjectives in SIMPLE: Theoretical Observations. *Proc. of the 2nd International Conference on Language Resources & Evaluation.*

Russian Associative Dictionary (2002). Ju.N.Karaulov et al. Moscow.

Sharoff, Serge. (2001). The Frequency Dictionary for Russian. http://www.artint.ru/projects/frqlist.asp

Soler, Clara. (2004). Extension of the SpanishWordNet. *Proc. of the 2nd Global WordNet Conference.*

Vasilevich A.P. et al. (2008). Colour and colour names in Russian. Moscow: URSS.

Wetzer, Harrie. (1996) The typology of adjectival predication. Berlin: Mouton de Gruyter.

Yavorskaya, Maria and Azarova, Irina (2009) Structuring of attributive word meanings in RussNet (in the group of Russian perceptional adjectives). *Proc. of the 15th International "Dialog" Conference*, Moscow. http://www.dialog-21.ru/dialog2009/materials/html/87.html

Zholkovskij, A., Mel'chuk, A. (1965). O vozmozhnom metode i instrumentax semanticheskogo sinteza. (*On a Possible Method and Instruments for Semantic Synthesis.*) Nauchno-texnicheskaja informacija. http://www.neuvel.net/meaningtext.html