"What Constructional Profiles Reveal about Synonymy: A Case Study of Russian Words for SADNESS and HAPPINESS"

Abstract

We test two hypotheses relevant to the form-meaning relationship and offer a methodological contribution to the empirical study of near-synonymy within the framework of cognitive linguistics. In addition, we challenge implicit assumptions about the nature of the paradigm, which we show is skewed in favor of a few forms that are prototypical for a given lexical item. If one accepts the claim of construction grammar that the construction is the relevant unit of linguistic analysis, then we should expect to find a relationship between the meanings of words and the constructions they are found in. One way to investigate this expectation is by examining the meaning of constructions on the basis of their lexical profile; this line of research is pursued in collostructional analyses. We have taken a different approach, examining the meaning of near-synonyms on the basis of what we call their "constructional profile". We define a constructional profile as the frequency distribution of the constructions that a word appears in. Constructional profiles for Russian nouns denoting SADNESS and HAPPINESS are presented, based upon corpus data, and analyzed quantitatively (using chi square and hierarchical cluster analysis). The findings are compared to the introspective analyses offered in synonym dictionaries.

Key words: synonymy, construction grammar, corpus data, Russian

1. Introduction

There are many ways to investigate the relationship between form and meaning. This study explores the relationship between the meaning of a noun and both the range and frequency of constructions that a noun appears in. We introduce the term "constructional profile" to describe the distribution of constructions associated with a given noun. There are two hypotheses: 1) Each noun will have a unique constructional profile, and 2) Similarity of meaning is correlated with similarity of constructional profile. The second hypothesis entails the expectation that closer synonyms will have constructional profiles that are more similar than synonyms that are further apart, and

synonyms will have more similar constructional profiles than semantically unrelated words.

The grammar of a case language (in which case is obligatory and morphologically marked in noun phrases) facilitates the operationalization of the hypotheses by providing objective measures to distinguish constructions, namely case markings and prepositions. Corpus data can be used to determine the distribution of constructions, and quantitative techniques can be applied to analyze these measures. Thus this study fulfills the criteria for "state of the art" corpus-oriented usage-based linguistics (Geeraerts 2005, Tummers et al. 2005).

Both the focus on morphology and the focus on the noun phrase set the present study apart from most work that has been done on meaning and constructions. Feldman (2006: 260-1, 318) notes that, aside from intonation and gesture, there are three mechanisms for conveying semantic relations in language, and they are: 1) words, 2) word order, 3) word form (inflection). Feldman points out that work in linguistics has been preponderantly based on English, and "this helps explain why much less attention has been paid to morphology as a source of meaning than to words and word order". Feldman's comment about the bias toward non-inflected languages is applicable to work in construction grammar, though there are some notable exceptions (Barðdal 1999 & 2006, Fried 2005). While Goldberg's work involves verb phrases and indeed most work in construction grammar is restricted to verb phrases, Goldberg (2006: 5, 221) acknowledges that "[a]ll levels of grammatical analysis involve constructions" and that constructions can profile units other than verbs. By using constructional profiles to probe the behavior of synonyms, this study also departs from the tradition of relying on lexical collocations to examine related meanings (cf. Kilgarriff 1997, Ide & Véronis 1998, Kobricov 2004, and Budanitsky et al. 2006).

We use emotion terms to test the "constructional profile" method for a number of reasons. One reason is that emotion terms are abstract and thus should be less prone to select the constructions they appear in based on ontological types. A concrete noun denoting a SURFACE, for example, would be predisposed to occur in constructions for 'onto', 'on' and 'off of'[1]. Emotion terms lack direct physical correlates that would limit the data in this way, presenting more complex constructional profiles. Secondly,

---

[1] The constructional profile of the SURFACE noun *stul* 'chair' suggests that it does indeed occur predominantly in precisely these constructions, cf. Section 4.3.

emotion words have traditionally been a focus of attention in both cognitive linguistics and Slavic linguistics (Apresjan 1993, Dziwirek forthcoming, Dziwirek & Lewandowska-Tomaszczyk 2003, Kövecses 2001, Lakoff & Johnson 1980, Levontina & Zalizniak 2001, Mostovaja 1998, Radden 1998, Wierzbicka 1998 & 1999), primarily due to their metaphorical nature and the various ways they are understood in different languages. Finally, there is some disagreement among synonym dictionaries (Abramov 1994, Aleksandrova 1998, Apresjan et al. 1997, Evgen'eva 2001, Švedova 2003) concerning the grouping of the Russian 'sadness' terms. The various proposals in these sources constitute hypotheses that can be tested using constructional profiles as a measure.

Russian lacks an umbrella term that would be equivalent to the English word *sadness,* relying instead on a series of synonyms: *grust', melanxolija, pečal', toska, unynie,* and *xandra.* There are clearly differences among the meanings of these words for 'sadness', since it is possible for native speakers to produce sentences like this one:

> *Uxodiš', i ja gljažu vsled tebe s grust'ju, no bez toski.*[2]
>
> [Depart, and I-Nom look following you-Dat with sadness-Inst, but without sadness-Gen.]
>
> 'You leave and I watch you go with sadness$_{grust'}$, but without sadness$_{toska}$.'

As this example suggests, SADNESS$_{grust'}$ is the kind of sadness that is associated with grief, whereas SADNESS$_{toska}$ is the sadness associated with yearning. However, such contrastive examples are rare, and conclusions of this sort are subjective and introspective. This study uses constructional profiles of the Russian SADNESS terms as an objective measure to probe the relationships among synonyms. The behavior of the SADNESS terms is further compared with that of a series of antonyms denoting HAPPINESS: *likovanie, naslaždenie, radost', udovol'stvie,* and *vostorg.*

This article is organized as follows. Section 2 gives an overview of relevant scholarly contributions to synonymy and the relatedness of word meanings, both outside of and within the framework of cognitive linguistics. Section 3 addresses the theoretical assumptions made on the basis of construction grammar and defines the term "constructional profile". The methodology is presented in Section 4, along with illustrative examples of constructional profiles. Section 5 undertakes the analysis of

---

[2] A. A. Bestužev-Marlinskij. *On byl ubit.* (1835-1836). All examples are cited from the Russian National Corpus (www.ruscorpora.ru).

the constructional profiles for the Russian SADNESS terms, and compares them to those of the HAPPINESS antonyms. Conclusions and possible additional uses for constructional profiles are offered in Section 6, which is followed by an Appendix presenting the data used in the quantitative analyses.

2. Approaches to Synonymy and Relatedness of Meanings

Of course the idea that a word's use is indicative of its meaning is not new, since it can be traced to the works of many linguists (among them Meillet, Bloomfield, and Harris, cf. Ide & Véronis 1998: 23 and others cited below). This section presents a brief overview of relevant scholarly works synonymy and related problems (polysemy, acquisition of word meanings). For the sake of organization, the discussion is broken down into three sections representing different linguistic approaches, although there is some overlap among them. This discussion is intended to be representative rather than exhaustive.

2.1. Computational Linguistics

Computational linguists have developed an impressive array of programs designed to detect and even "learn" how to disambiguate polysemous words and recognize synonyms (two closely related problems in Word Sense Disambiguation, WSD). The majority of work in WSD has followed Firth's (1957: 11) maxim "You shall know a word by the company it keeps", thus focusing on word co-occurrence data to determine word senses and their relative "distance". Such algorithms typically look at a node word and the window of x words (for example, if x=3, the window would include three words to the left and three to the right) that surround it in all its occurrences in a corpus and then compare this measure to that of other words. In most WSD studies, grammatical information (syntax, morphology, word order) is not taken into account, although there is some indication that the algorithms are sensitive to grammatical facts such as word class (Burgess et al. 1998). Where grammatical information has been included, this has typically been limited to identification of part of speech (Ide & Véronis 1998: 20). Latent Semantic Analysis (LSA; Landauer et al. 1998) computes the aggregate of all the contexts that a given word appears and does not appear in and represents this as a high-dimensional "semantic space". LSA is a significant improvement over many earlier methods which relied on dictionaries and manually-crafted semantic networks; its only input is raw text parsed into words and

passages. Though LSA can mimic some human activities (synonymy and word relatedness tests), it also makes some odd judgments (e.g., that English *verbally* and *sadomasochism* are closely related). Burgess et al. (1998) offer the Hyperspace Analogue to Language (HAL), which likewise computes the statistical co-occurrence patterns of words, depicted as multidimensional scaling solutions, and one of the conclusions is that synonyms do indeed occur in the same contexts. Dagan (2000) offers another solution, one that features semi-automatic thesaurus construction procedures based on corpus statistics. Like LSA and HAL, Dagan's model is based entirely on word co-occurrence vectors; it is claimed that there is no adequate parser that would make it possible to include grammatical information (Dagan 2000: 462). Turney (2002, 2005) has worked on two further options, Pointwise Mutual Information (PMI) and Latent Relational Analysis (LRA), which have been tested against items from the TOEFL and SAT tests, respectively. Though the LRA has a grammatical component, both are based primarily on word co-occurrence data. Kilgarriff (1997), Ide & Véronis (1998), Kobricov (2004), and Budanitsky et al. (2006) provide overviews of the trajectory of research in WSD. Their conclusions are rather disappointing, despite the variety and computational sophistication of the models devised. Nouns in particular have been most resistant to WSD (Ide & Véronis 1998: 21). Both Kilgarriff (1997) and Ide & Véronis (1998) make the point that despite computational advances, one of the crucial theoretical issues, namely defining what a word sense is, remains unresolved, and this has severely hampered progress, particularly since the underlying assumption is usually that word senses are discrete and independent of corpora. Kilgarriff (1997) and Budanitsky et al. (2006) cite work on polysemy and metaphor (specifically Apresjan 1974, Lakoff & Johnson 1980 and Lakoff 1987), arguing that if word senses do not behave as classical categories, and can also be influenced by ad-hoc categories, then it is very unclear how such "semantic relationships could be quantified in any meaningful way, let alone compared with prior quantifications of the classical and non-classical relationships" (Budanitsky et al. 2006: 45). Ide & Véronis (1998: 27) conclude that "relatively little progress seems to have been made in nearly 50 years" and "it appears that we may have nearly reached the limit of what can be achieved in the current framework". Kobricov's (2004) evaluation is nearly identical, stating that even when good accuracy has been achieved, it applies only to a very small group of words, and that the best descriptor for the state of the art is "stagnation".

In sum, computational approaches to synonymy and polysemy are invested primarily in observing lexical collocations, largely to the exclusion of syntactic information. These approaches have achieved limited results, and have proved least useful in work on nouns.

2.2. Bootstrapping and Frames

Bootstrapping approaches (Gleitman & Gillette 1995, Lidz et al. 2001) are concerned with how the meanings of words are acquired and stored and what role syntactic information plays in this process. At issue is the fact that verbs are relatively abstract: you often can't point to a real-world action, and the uses of verbs are often asynchronous with corresponding actions (Gleitman & Gillette 1995: 415). Furthermore, many verbs are synonymous, and speakers are able to distinguish among near-synonyms. The hypothesis is that syntactic range information makes it possible for learners to fix the meaning of novel verbs. Syntactic range information specifies what types of constructions a verb typically appears in, without reference to relative frequency. Two series of psycholinguistic experiments (Gleitman & Gillette 1995, Lidz et al. 2001) support this hypothesis, with evidence that both children and adults use systematic structural information in order to interpret English verbs. Dąbrowska (forthcoming) further argues, on the basis of experiments with English verbs of walking and running, that syntactic range information is supplemented by speaker's knowledge of collocational patterns in distinguishing the meanings of close synonyms.

The lexicographic research that serves as the basis for FrameNet has developed a sophisticated means of analyzing semantic frames, linking "the meaning of words very explicitly to the syntactic contexts in which those words occur" (Atkins et al. 2003: 253). This approach, like bootstrapping, focuses on identifying the range of syntactic constructions in which a word occurs, in addition to the collocational preferences. Frame elements focus mostly on the behavior of verbs and can yield subtle analyses of synonyms.

2.3. Behavioral Profiles and Collostructions

Karlsson (1985, 1986) observed, on the basis of Finnish data, that in a language with complex inflectional morphology, the majority of forms in a given paradigm are unattested or of very low frequency in a corpus. Most paradigms are instantiated by a

fairly small number of stereotypic forms that are the "morphological analogues of the prototypes in Rosch's theory of word meaning" (Karlsson 1985: 150). This observation has important implications, since most theories (cf. Karlsson's overview 1985: 137) assume that a paradigm is normally fully populated, and rule-based theories assume that all forms in a paradigm are generated by rules in an equiprobable fashion. The skewed frequency profiles found in Finnish led Karlsson (1986: 28) to assert that speakers probably use a combination of lexical storage and "rules" in relation to paradigms, a conclusion that comports well with the basic tenets of cognitive linguistics (cf. Dąbrowska 2004: 7-27; Croft & Cruse 2004: 291-327). Though Karlsson (1986: 27) does claim that the Finnish data show "how meaning properties are reflected in the use of forms", his conclusions are restricted to differences among broad classes of words. According to Karlsson, for example, mass nouns, count nouns, and proper nouns behave differently from each other as groups, but verbs are fairly homogeneous. Whereas Karlsson stopped short of implying that frequency profiles might provide finer-grained distinctions within types of nouns or synonyms, Arppe, also working on Finnish data (but with a larger corpus and more sophisticated software), has found that there are indeed differences among different types of mass nouns (Arppe 2001), and there are differences even among the near-synonyms meaning THINK (Arppe 2005).

Synonyms have been the focus of attention in the use of behavioral profiles (Atkins 1987, Hanks 1996), which can combine a variety of types of information, not limited to collocational and syntactic preferences. Geeraerts (1988) pioneered synonymy research in cognitive linguistics, comparing 19th century uses of two Dutch verbs meaning DESTROY. Geeraerts' study incorporates collocational, constructional, semantic and metaphorical data and uses corpus data to corroborate introspective analyses found in synonym dictionaries. Divjak & Gries (Divjak 2006, Divjak & Gries 2006 and Gries & Divjak forthcoming) tagged 87 variables (morphosyntactic, syntactic and semantic) in order to establish the behavioral profiles of Russian verbs meaning TRY and calculate the "distances" among near-synonyms. Glynn (forthcoming) applies a similar approach to investigate the semantic relationships within the polysemy of a single word ("parasynonyms" of English *hassle*), tagging corpus examples and performing a quantitative analysis. Collostructional analysis (Stefanowitsch & Gries 2003, 2005) takes the construction as the point of departure, investigating the range and frequency of words that appear in the construction. A

related strategy is metaphorical pattern analysis (Stefanowitsch 2006 a&b, Svanlund 2007), which can compare the metaphorical uses (based on the constructions) that near-synonyms appear in.

2.4. Relationship of Constructional Profiles to Previous Research

Most previous studies of synonyms focus on verbs, whereas the present study examines nouns. Unlike the approaches undertaken in computational linguistics, we define comparisons syntactically, in terms of constructions, instead of lexically, in terms of collocated words. Our approach can be understood as a reversal of the perspective of collostructional analysis, an option that has been proposed for future research (Stefanowitsch & Gries 2003: 237), but not yet pursued. This fresh perspective is facilitated by the fact that Russian is a case-marking language, making it possible to collect data with no subjective tagging component, based on the objective presence of morphological features. Bootstrapping and frame approaches focus on the range of syntactic contexts that a word appears in; our study additionally presents the frequencies of occurrence for relevant syntactic contexts.

3. Construction Grammar

We use the term *construction* in a way that is compatible with current usage in cognitive linguistics, in other words as used by Langacker (1987, 1990, 1991), Croft (2001), Goldberg (1995 and 2006), and Fillmore (Fillmore 1985, Kay & Fillmore 1999, Fillmore et al. forthcoming). Although some differences in the usage of *construction* among these scholars must be acknowledged (cf. Langacker 2003 and Goldberg 2006: 213-226), these points are less relevant to our analysis than the ideas that all three share, so we will focus on their common ground, ignoring minor discrepancies.

Our definition of construction is: "a conventionalized pairing of form and meaning in a language". This definition is closest in its phrasing to Goldberg's (2006: 3), yet consistent in spirit with Langacker's (1987: 58) "symbolic unit" which pairs form (phonological pole) with meaning (semantic pole).

Our constructions are of the form: "*[(preposition) [NOUN]_{case}]*". This formula states that case is obligatory in all constructions, but only some constructions also involve a preposition. This formula states that the noun elaborates (Langacker 1987: 68; 304-5) the construction that is schematically specified by the case and preposition

by filling the placeholder for the noun. Because the noun is the variable part of the construction, we often use a short-hand formula, stating the components "*case*" or "*preposition + case*". For each construction, this form is paired with a meaning that is only partially determined by the meanings of the components. The meaning of each construction is emergent (Langacker 1991: 5-6, 534; Bybee & Hopper 2001: 2, 10; MacWhinney 2001), motivated by the patterns of uses over the various nouns that appear in the construction, and also by the larger (clause-level) constructions that these noun phrase constructions appear in. Our analysis gives empirical substance to the claims made by Raxilina (2000) that Russian nouns can serve as constructional cores, and that the meaning of a noun is partly a function of the constructions it is found in. This analysis is also in harmony with the traditions of the Moscow semantic school (Apresjan 1995, Mel'čuk 2001) and the school of "Logical Analysis of Language" (Arutjunova 2007), which likewise assert that combinatorial properties of nouns reveal the cognitive structure of nominal semantics.

It is unlikely that speakers store all uses of given words and constructions, but there is evidence that people use generalizations about the frequency of word use (Goldberg 2006: 62, 46). These generalizations can serve as the basis for creating abstract schemas for constructions, establishing correlations between form and meaning. Goldberg (2006: 104-119) argues that constructions have strong associations with meaning by virtue of their advantages in terms of both cue validity and category validity. Cue validity refers to the likelihood that a given meaning will be present given the presence of a certain item. In a study comparing the cue validity of words (verbs) with constructions, Goldberg found that words and constructions have roughly equal cue validity, which means that knowing that a linguistic unit contains a given word gives you about the same predictive information as knowing that a linguistic unit occurs in a given construction. However, because there are far fewer constructions than lexical items in a language, constructions are far more available in terms of determining meaning. Category validity is the likelihood that a certain item will be present when the meaning is already given. In Goldberg's studies the category validity of constructions is found to be far higher than that of words (verbs). In other words, if you know that a unit expresses a certain meaning, it is much easier to predict what construction might be present than to predict what word the unit might contain. Goldberg has thus empirically established the connections between constructions, frequency and meaning. Although Goldberg's work focuses

on verbs as construction cores, we argue that her conclusions are applicable to noun phrases, particularly in languages that mark case.

The morphological marking on Russian nouns makes them much more information-rich in terms of specifying what construction is present than English nouns. Kempe & MacWhinney (1999) have established, based on psycholinguistic data, that case in Russian has high cue validity and high cue availability, even relative to another case-marking language (German), and that Russian speakers do rely on case in on-line sentence interpretation. On this basis we assert that Goldberg's claims for the relationships between constructions and verbs in English are applicable to nouns in Russian as well.

3.1. Constructional Profiles

A *constructional profile* is a property of a word. *Constructional profile* can be defined as: "the relative frequency distribution of constructions that a given word appears in". In other words, let us say that the word *LEXEME* can appear in constructions $C_1 \ldots C_n$. In order to arrive at *LEXEME*'s constructional profile, it is necessary to gather data on the frequency of *LEXEME*'s occurrence in each of the constructions $C_1 \ldots C_n$ and to compare those frequencies as percentages of *LEXEME*'s overall occurrence (a.k.a. the "reliance" metric, cf. Schmid 2000: 54). *LEXEME*'s constructional profile is thus a chart showing that *LEXEME* occurs X% of the time in construction C_1, Y% of the time in construction C_2, Z% of the time in construction C_3, etc. through C_n. Each percentage indicates how frequent the given construction is for the given word in a particular corpus, and the aggregate of percentages indicates the degree to which that noun is associated with that particular pattern. Constructional frequency data is extracted from corpora that are designed to reflect the parameters of a given language. In practice, there are often many constructions associated with a given word, and most occur at very low frequencies. Based on the data in our study, usually only 6-10 constructions are needed to accurately represent the constructional profile of a word.

Constructional profiles can be likened to flavors. Flavors are composite values of the variables that our tongue and nose can perceive (Churchland 1995). In other words, a flavor such as apricot is a collection of peaks with various values that differs from other collections of peaks such as the one associated with peach. A word's constructional profile is probably unique and representative of its meaning, though there are certainly other factors, such as: the embodied contexts in which a word is

used, the knowledge structures (frames; cf. Fillmore 1982) it is associated with, its collocational patterns, and transparent etymological or derivational relationships to other words in the lexicon.

It is tempting to consider a possible relationship between constructional profiles and entrenchment, given assumptions that have been made about increases in neural connections as a function of frequency (Langacker 1987: 59-60, 100, 380; Langacker 1991: 45, Bybee & Hopper 2001: 9; Taylor 2002: 276; Dąbrowska 2004: 213, 223; Feldman 2006: 105). This connection is expressed most explicitly in Schmid's (2000: 39) From-Corpus-To-Cognition Principle: "Frequency in text instantiates entrenchment in the cognitive system". Some recent work (Schmid 2007, Schmid forthcoming, Gilquin 2007a&b) has pointed out that corpus frequency may be an imperfect measure of entrenchment. Given these reservations, we remain agnostic and make no claims concerning a connection between constructional profiles and entrenchment.

Russian has six cases marked by means of synthetic inflectional endings: Nominative (Nom), Accusative (Acc), Dative (Dat), Instrumental (Inst), Genitive (Gen), and Locative (Loc). Since every noun phrase obligatorily expresses one of these cases[3], a noun phrase will always carry with it the syntactic and semantic information associated with the given case. The Russian cases present a complex system with dozens of submeanings. However, Janda and co-authors (Janda 1993, 1999, 2000, 2002a-d, 2004, forthcoming; Janda & Clancy 2002; Divjak & Janda forthcoming) have established that each Russian case forms a coherent semantic whole. In keeping with the above-cited research on Russian case, we will assume that case is the primary marker of the meaning of a syntactic relationship and that prepositions, where present, elaborate those meanings, forming a composite structure that shows conceptual integration (Langacker 2003). All of the Russian cases can appear with various prepositions and five of them can appear without a preposition. The various combinations of case with and without prepositions yield seventy potential constructions of the form *[(preposition) [NOUN]$_{case}$]* for any given noun in Russian (Janda & Clancy 2002).

---

[3] Some indeclinable nouns, such as *kino* 'cinema' constitute an exception to this rule, though all numerals, adjectives, determiners and pronouns that modify such nouns bear the appropriate case marking. This indicates that for such nouns case is present, though the entire paradigm may be syncretic.

4. Methodology

The corpora from which data were extracted and the methodology used in the process are described below, illustrated by an example of the results and how they are presented in this article.

4.1. Corpora

Our study extracted data from two corpora, the Russian National Corpus (http://www.ruscorpora.ru; henceforth RNC) with over 120 million words, and the Biblioteka Maksima Moškova (http://lib.ru/; henceforth BMM) with over 600 million words. Both corpora consist exclusively of authentic texts produced by and for native speakers, and their contents have been edited for typographical accuracy. One major difference between the two corpora is that the RNC has been designed to reflect a greater range of genres, including samples of popular written and spoken Russian, whereas the BMM is literally an electronic library of primarily literary works.

A pilot study was conducted to determine how many sentences would be needed for constructional profiles and to compare results across and within corpora. That study indicated that 500 sentences for each noun would yield sufficient results that were reliably stable for the corpora (the vast majority of data differed by a fraction of a percentage point or less, with the ceiling of differences at about two percentage points).

4.2 Data extraction

500 sample sentences were extracted for each word in the study in order to determine each noun's constructional profile. The pertinent noun phrase construction (the one containing the queried word) in each sentence was analyzed manually, and the case of every queried noun was recorded along with the identity of any associated preposition. The analyses were conducted by students in the Linguistics Department at the University of Kazan'. In principle it would be possible to have the analysis done by machine, but there is no automatic parser of Russian at present with sufficient accuracy. This does not mean, however, that any subjective judgments were involved. On the contrary, the identity of the case a noun appears in is unambiguous in Russian despite a small amount of paradigmatic syncretism. In rare instances where there might be some confusion, a native speaker can easily recover the case by asking the

relevant Who?/What? question, which gives a unique answer for each case (*Kto?/Čto?* for Nominative, *Kogo?/Čto?* for Accusative, etc.). The task was simple and objective, the analyses were carried out by linguists in training on their native language, and the results were virtually error-free.

Once the data was collected and analyzed, the scope of the search could be narrowed down to target the most valuable results. A given noun can usually appear in a fairly large number of constructions, but most of these are of such low frequency (<1%) that they contribute little information about the noun's overall constructional profile. On the other hand, there are some constructions that appear in fairly high frequencies (e.g., Nominative subject, adnominal Genitive), but do not give much information about aspects of a noun's constructional profile because virtually any noun can appear in those constructions. The search was thus narrowed to those sentences in which the noun appeared as a non-subject argument of the verb or as an adverbial. This made it possible to focus on the noun phrase constructions that were most relevant to the verb. Thus from the original 500 sentences, only the sentences where the queried word appeared as a non-Nominative argument or adverbial were analyzed further, and all frequencies are based upon the remaining number of sentences for a given word (usually in excess of 70% of the original 500). From these data it is possible to pinpoint which constructions are most representative of a word's constructional profile to present their frequencies. The Direct Object construction is fairly frequent for most nouns, and data on that construction is included in calculations, but not in graphs, to highlight the constructions that are most relevant.

4.3 Sample analysis

Graph 1 gives the constructional profiles of one SADNESS noun, *pečal'* and three non-synonymous nouns: *stul* 'chair', *utka* 'duck', and *mečta* 'dream' (data is presented in tables in the Appendix).

Graph 1: Constructional Profiles of Non-Synonyms
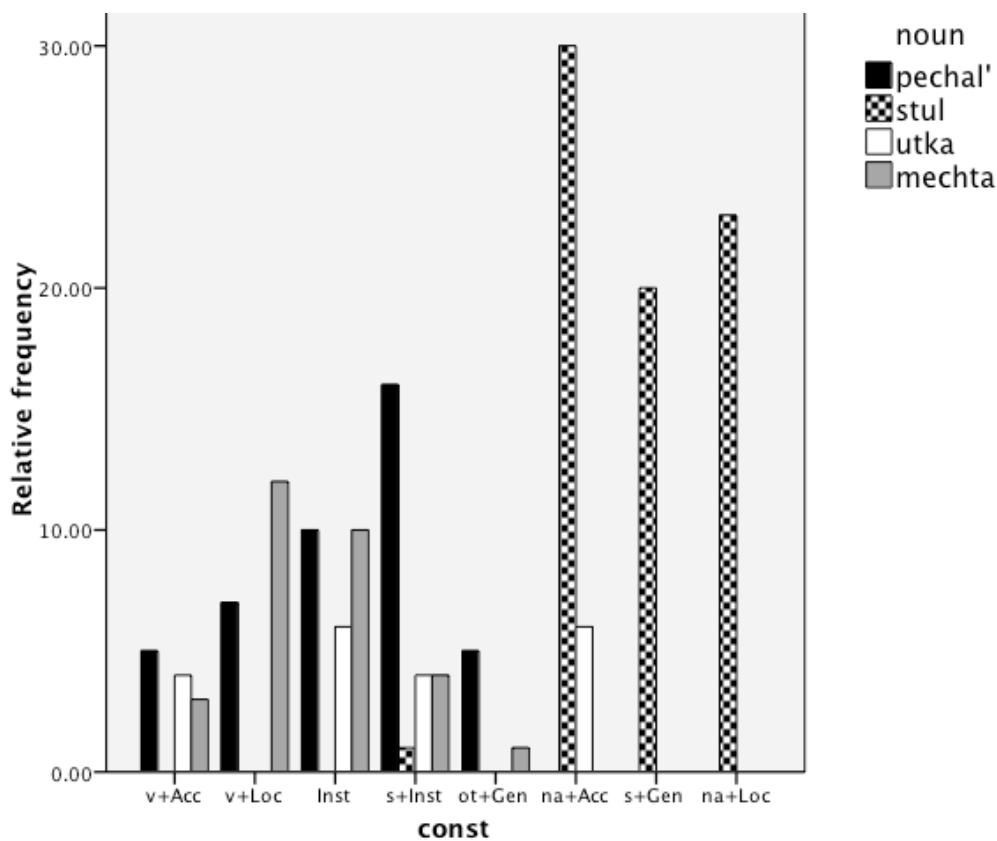
We see that the both the range and relative frequencies of the constructions associated with these nouns differ. The relevant constructions can be paraphrased as follows: *v*+Acc 'into/at', *v*+Loc 'in(side)', Inst 'by means of/as', *s*+Inst 'with', *ot*+Gen '(away) from', *na*+Acc 'onto', *s*+Gen 'off of', *na*+Loc 'on'. The two abstract nouns, *pečal'* and *mečta*, share the range of constructions, appearing in the first five constructions (though in different frequencies), but not in the last three. *Utka* 'duck' is found in the *v*+Acc, Inst, *s*+Inst, and *na*+Acc constructions. *Stul* 'chair' is dominated by *na*+Acc, *s*+Gen and *na*+Loc. The chi square value of 1014.8 is highly significant (p < 0.0001 for df=27), indicating that these differences cannot be attributed to chance. Furthermore, the Cramer's V (indicating the strength of the chi square effect) is 0.495, which qualifies as a large effect (cf. King & Minium 2008).

5. Case Studies

The constructional profiles of the Russian nouns for SADNESS and HAPPINESS are presented in graphs and subjected to statistical analyses. Chi square results ensure that the effects are not the result of chance and hierarchical cluster analysis measures the "distances" between words, indicating which near-synonyms are closer and which are farther apart. The latter results can be used to corroborate the groupings found in synonym dictionaries.

5.1. Russian nouns for SADNESS

Russian synonym dictionaries struggle with this set of nouns. Most often, *pečal'*, *toska*, and *grust'* are placed in one group, characterized as denoting the 'unpleasant feeling one has when one wants something one doesn't have and doesn't believe one can get it' (Apresjan et al. 1997). *Melanxolija* and *xandra* are listed as another group, and then there is disagreement over what to do with *unynie*. Apresjan et al. 1997 groups *unynie* with *pečal'*, Aleksandrovna 1989 puts *unynie* with *xandra*, and Evgen'eva 2001 puts *unynie* with both *grust'* and *xandra*, claiming that it has two meanings. Švedova 2003 unites *unynie* with *grust'*, *xandra* and *melanxolija*.

The constructional profiles of these words both confirm the overall pattern suggested in synonym dictionaries and explain why there is a problem with *unynie*. Graph 2 shows the constructional profiles for the SADNESS nouns.

Graph 2: Constructional Profiles of SADNESS nouns
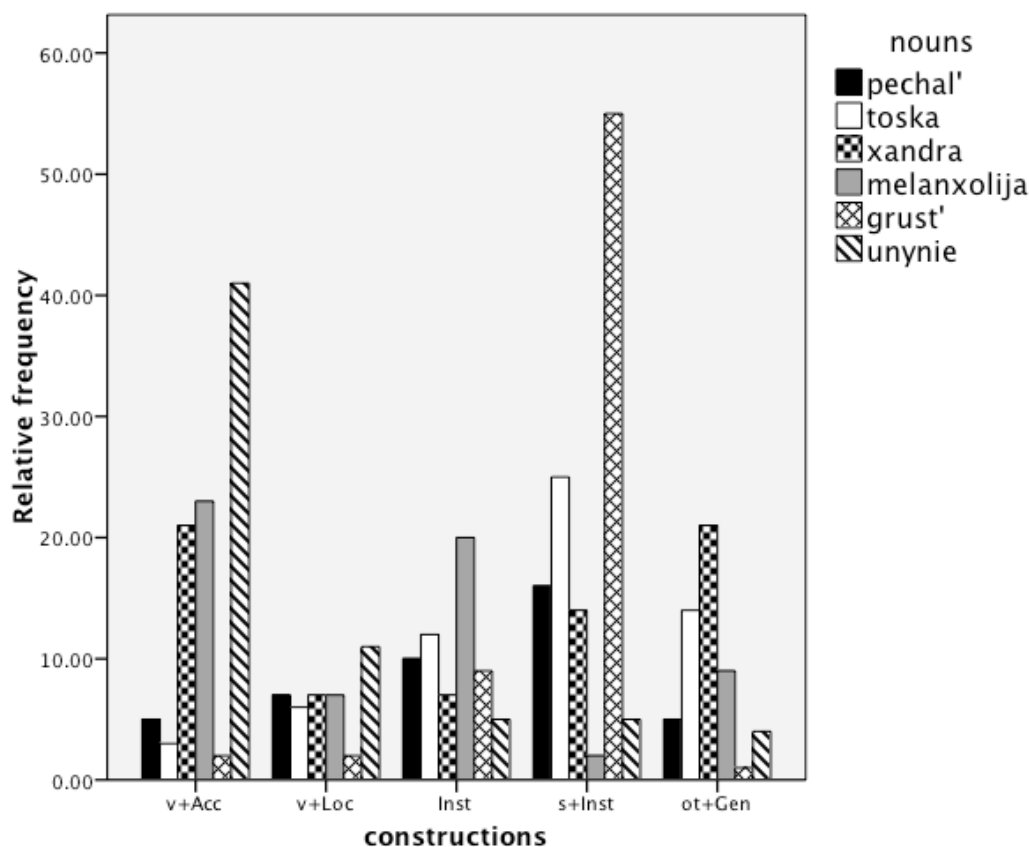
The relevant constructions are: *v*+Acc 'into/at', *v*+Loc 'in(side)', Inst 'by means of/as', *s*+Inst 'with', *ot*+Gen '(away) from'. The constructional profiles provide a variety of information on the behavior of the SADNESS synonyms. To begin with, these

six nouns all show the same range of constructions in their profiles, which was not the case for the non-synonymous nouns in Graph 1. Though the range is shared, the distributions within this range are significantly different, as demonstrated by the chi square value which is 730.35, and the Cramer's V of 0.305 which qualifies as a moderate effect ($p<0.0001$, df=30 for both values). If we compare these results to the results for the non-synonyms, we see that though nouns in both groups are significantly different from each other, the chi square and effect values are greater for the non-synonyms than for the synonyms.

Next we notice patterns within the data: *pečal'* and *toska* have similar values for the first three constructions, but dissimilar ones for the last two. *Xandra* and *melanxolija* have similar values for the first two constructions, but dissimilar values for the last two. *Grust'* and *unynie* look like outliers: *grust'* is dominated by the *s*+Inst 'with' construction, whereas *unynie* gives the highest values in the group for the *v*+Acc 'into/at' and *v*+Loc 'in(side)' constructions.

The suggestion that some near-synonyms are closer to each other than others can be tested mathematically, using hierarchical cluster analysis to measure this pheonomenon in terms of squared Euclidian distances (cf. proximity table in the Appendix). By this metric, the closest SADNESS synonyms are *pečal'* and *toska* (separated by 5.844), the next closest item is *xandra* (7.968), followed closely by *melanxolija* (8.041). *Grust'* (11.705) joins the group next, followed by *unynie* (12.798). These proximity values yield the hierarchical cluster in Graph 3:

Graph 3: Hierarchical Cluster of SADNESS nouns

| pečal' | toska | xandra | melanxolija | grust' | unynie |
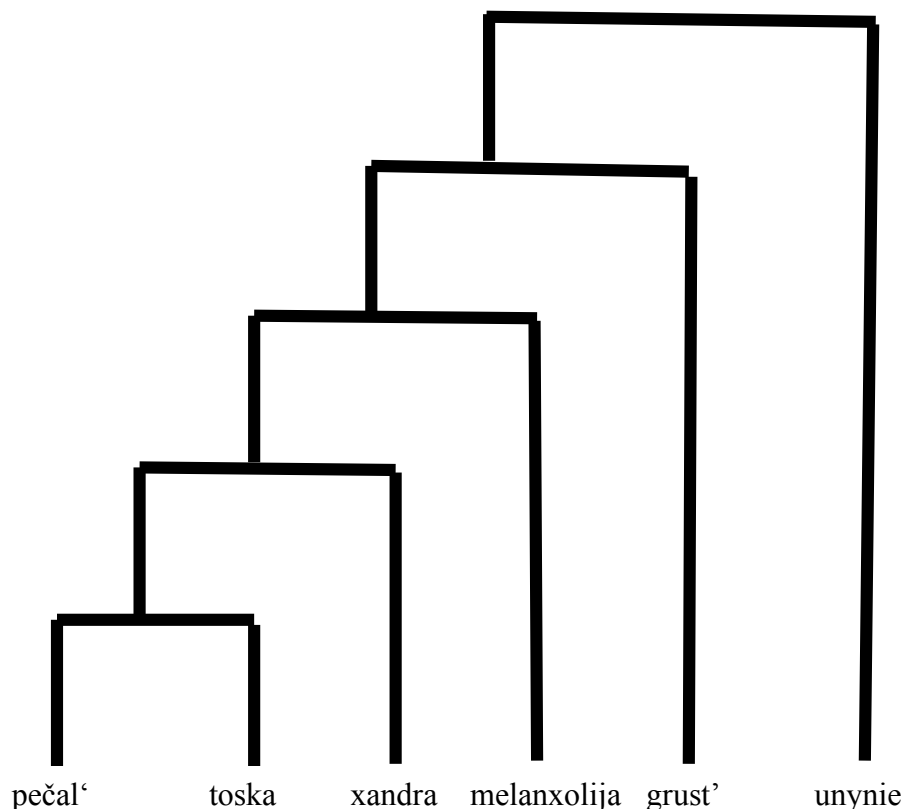
In comparing the constructional profiles with the groupings in synonym dictionaries, we see that the grouping of *pečal'* with *toska* and *melanxolija* with *xandra* is well-justified. The difficulty with *unynie* is unsurprising, since it is indeed the most extreme outlier in the group.

5.2. Russian nouns for HAPPINESS

Antonyms are words that are virtually identical to each other in terms of what domain they refer to and what they profile within that domain, but have opposite values for some part of their meaning (Croft & Cruse 2004: 164-192). Both SADNESS and HAPPINESS are states involving human emotions evaluated on a scale of wellbeing, so their meanings are in many ways similar. In Russian, it turns out that the same set of constructions is most relevant for both groups of synonyms, making it easy to compare these groups of nouns.

Synonym dictionaries are less clear in making distinctions among these nouns. Where distinctions are made, it appears that *vostorg* is treated as the outlier: Aleksandrova (1998) defines all the other happiness nouns in terms of each other

while setting *vostorg* apart, whereas Švedova 2003 places *naslaždenie*, *radost'*, and *udovol'stvie* in one group and *likovanie* and *vostorg* in another.

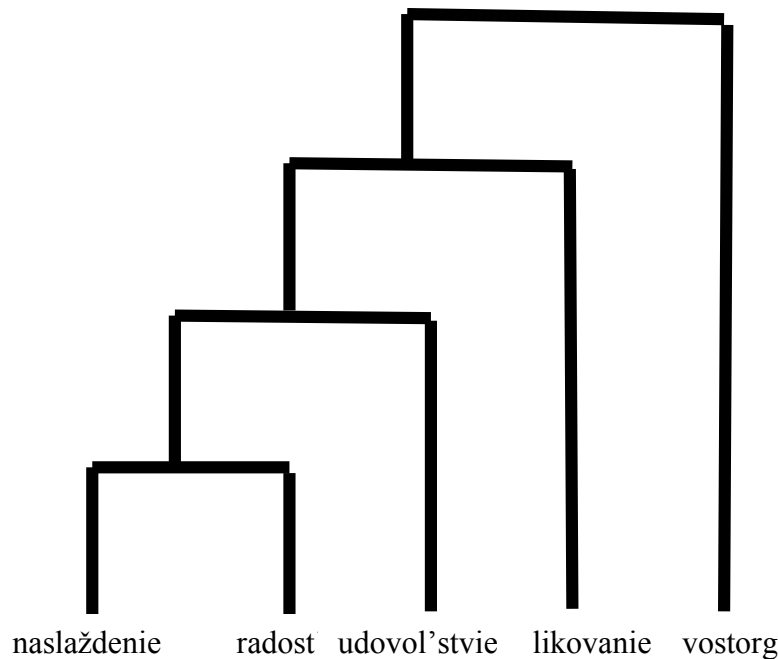Graph 4 presents the constructional profiles of the HAPPINESS nouns.

Graph 4: Constructional Profiles of HAPPINESS nouns
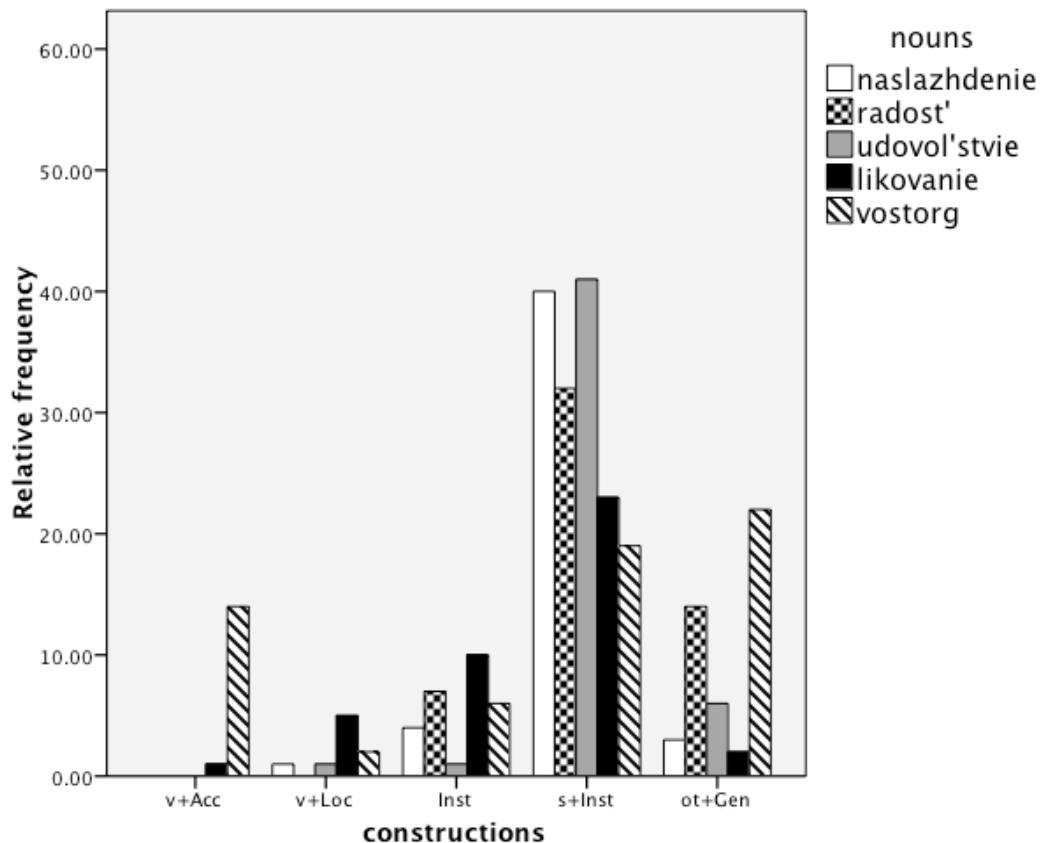
The chi square value of 469.4, Cramer's V of 0.264 (p<0.0001, df=24) indicate that these nouns are significantly different from each other, though the effect is slightly less than for the SADNESS nouns. The constructional profiles corroborate the patterns in the synonym dictionaries. *Naslaždenie*, *radost'*, and *udovol'stvie* do indeed pattern similarly, with zero or low values for the first three constructions, a peak at the fourth and lower values again for the last construction. *Vostorg* behaves like an outlier, with high values for the first, fourth, and fifth construction. *Likovanie* appears to fall somewhere between the first group and *vostorg*, for it is the only other noun with a non-zero value for the first construction, has its peak with the fourth construction and then a low value for the last one. This grouping is also confirmed by the hierarchical cluster analysis (cf. proximity table in Appendix), which finds *naslaždenie* and

radost' as the closest synonyms (separated by 3.512), closely followed by *udovol'stvie* (3.979). Further out lie *likovanie* (9.632) and finally *vostorg* (13.22). The proximity values yield the hierarchical cluster in Graph 5.

Graph 5: Herarchical Cluster of HAPPINESS nouns

naslaždenie   radost   udovol'stvie   likovanie   vostorg

Once again, the constructional profiles largely confirm the suggestions made by synonym dictionaries while pinpointing the source of disagreements among them.

6. Conclusions

We present the constructional profile, the relative frequencies of constructions a word appears in, as a possible measure of a word's meaning. The constructional profile patterns of synonyms are shown to share a small group of constructions that they appear in most frequently. Differences in frequencies correspond to differences in "distance" between synonyms. Constructional profiles largely confirm the introspective judgments of dictionary authors, and in addition pinpoint where the differences among synonyms lie. Antonyms largely share the set of constructions they appear in and may show overlap in constructional profile patterns. Unrelated words share neither property. Constructional profiles provide an opportunity for empirical verification of hypotheses relevant to a usage-based approach to linguistics.

Constructional profiles may have potential use in exploring the metaphorical behavior of words, thus building upon current work on metaphorical pattern analysis (Stefanowitsch 2006 a&b, Svanlund 2007). It would be possible to compare the constructional profiles of concrete source domain nouns and corresponding nouns in a metaphorical target domain. The CONTAINER metaphor is often cited as relevant for the domain of emotions (Lakoff & Johnson 1980: 31-2, Kövecses 2001: 37), and constructions with verbs involving 'entering into' (usually by means of 'falling' or 'sinking into') an emotional state have been associated with SADNESS nouns in both Russian and Polish (Wierzbicka 1998: 11). It would be possible to test this connection empirically by finding the constructional profiles of a group of CONTAINER nouns and comparing them to the constructional profiles of emotion terms. The data here suggest that different nouns may behave differently in terms of their metaphorical extensions. Two constructions relevant for CONTAINERS, *v*+Acc 'into/at' and *v*+Loc 'in(side)' are more prominent among the SADNESS nouns, particularly *unynie*, followed by *melanxolija* and *xandra*, than among the HAPPINESS nouns, where they are relevant for *vostorg*. Indeed, these appear to be the emotions in Russian that one can get into or be in. Curiously, the corresponding construction for leaving a CONTAINER, namely *iz*+Gen 'out of', is absent from the constructional profiles of the SADNESS and HAPPINESS nouns (though isolated examples can be found in a corpus). The only conventional means for departing these emotional states seems to involve a DISEASE metaphor using the *ot*+Gen '(away) from' construction, as in this example:

> *Samoe lučšee lekarstvo <u>ot xandry</u> èto čtenie.*[4]
>
> [The best medicine-Nom <u>from sadness-Gen</u> that reading-Nom.]
>
> 'The best cure <u>for sadness</u> is reading.'

The observation that departing a state of SADNESS seems to invoke a DISEASE metaphor is something that might be tested empirically by comparing the constructional profiles of some typical DISEASE nouns with those of emotion terms. Another use of the *ot*+Gen '(away) from' construction often interprets the emotion as a metaphorical CAUSE, as in this example:

> *Podumajte, ètot čelovek umer <u>ot melanxolii</u>!*[5]
>
> [Think, that person-Nom died <u>from sadness-Gen</u>!]
>
> 'Just imagine, that person died <u>of sadness</u>!'

---

[4] Paneva, A. Ja. *Vospominanija* (1889-1890).
[5] Lidija Ginzburg. *Zapisnye knižki. Vospominanija. Èsse.* (1920-1943).

Again, this obervation could be tested empirically.

Other potential uses for constructional profiles involve language acquisition and the relationship between storage and rules in a usage-based grammar. Constructional profiles suggest that certain forms in a paradigm are more prototypical for a given word than others. These patterns might correspond to order of acquisition among children and strategies for online use and interpretation among adults. Psycholinguistic experiments could test whether such correlations exist.

In sum, constructional profiles may prove to be a valuable metric for determining the relationship between meaning and use, and this metric may be used for a variety of investigations relevant to the usage-based model of cognitive linguistics.

Appendix

The following three tables give both the raw and relative frequencies used in all the charts and calculations. "DO" stands for the Direct Object construction, and "other" stands for an aggregate of all other constructions.

Unlike nouns

	pečal' 'sadness'		*stul* 'chair'		*utka* 'duck'		*mečta* 'dream'	
v+Acc	16	5%	1	0%	16	4%	11	3%
v+Loc	22	7%	0	0%	0	0%	42	12%
Inst	32	10%	2	0%	23	6%	36	10%
s+Inst	49	16%	3	1%	15	4%	15	4%
ot+Gen	16	5%	2	0%	0	0%	4	1%
na+Acc	0	0%	108	30%	23	6%	0	0%
s+Gen	0	0%	70	20%	0	0%	0	0%
na+Loc	0	0%	82	23%	0	0%	0	0%
DO	128	41%	64	18%	246	66%	174	50%
other	52	17%	25	7%	49	13%	63	18%
Total	315	100%	358	100%	372	100%	345	100%

SADNESS nouns

	pečal'		*toska*		*xandra*		*melanxolija*		*grust'*		*unynie*	
v+Acc	16	5%	8	3%	30	21%	52	23%	6	2%	126	41%
v+Loc	22	7%	16	6%	10	7%	16	7%	6	2%	33	11%
Inst	32	10%	33	12%	10	7%	45	20%	27	9%	16	5%
s+Inst	49	16%	70	25%	19	14%	5	2%	160	55%	16	5%
ot+Gen	16	5%	39	14%	29	21%	20	9%	3	1%	14	4%
DO	128	41%	84	30%	20	14%	32	14%	50	17%	25	8%
other	52	17%	33	12%	22	16%	57	25%	38	13%	82	27%
Total	315	100%	283	100%	140	100%	227	100%	290	100%	304	100%

HAPPINESS nouns

	naslaždenie		radost'		udovol'stvie		likovanie		vostorg	
v+Acc	0	0%	0	0%	0	0%	4	1%	42	14%
v+Loc	3	1%	1	0%	4	1%	21	5%	35	12%
Inst	12	4%	21	7%	4	1%	39	10%	18	6%
s+Inst	117	40%	92	32%	165	41%	90	23%	58	19%
ot+Gen	9	3%	41	14%	24	6%	8	2%	66	22%
DO	111	38%	106	37%	171	42%	171	44%	67	22%
other	39	13%	24	8%	39	10%	59	15%	18	6%
Total	291	100%	285	100%	407	100%	392	100%	304	100%

The following two tables are proximity matrices stating the squared Euclidian distances that establish the hierarchical clusters. The relevant values are bold-faced.

SADNESS nouns

Case	Squared Euclidean Distance					
	1:grust'	2:melanx	3:pechal'	4:toska	5:unynie	6:xandra
1:grust'	0.000	14.235	11.705	12.762	27.415	13.662
2:melanx	14.235	0.000	8.041	8.226	12.798	11.715
3:pechal'	**11.705**	**8.041**	0.000	5.844	17.123	14.679
4:toska	12.762	8.226	**5.844**	0.000	23.880	7.968
5:unynie	27.415	**12.798**	17.123	23.880	0.000	19.949
6:xandra	13.662	11.715	14.679	**7.968**	19.949	0.000

HAPPINESS nouns

Case	Squared Euclidean Distance				
	1:likovani	2:naslazd	3:radost'	4:udovol's	5:vostorg
1:likovani	0.000	9.632	12.526	14.144	25.993
2:naslazd	**9.632**	0.000	3.512	3.979	20.455
3:radost'	12.526	**3.512**	0.000	8.550	13.220
4:udovol's	14.144	**3.979**	8.550	0.000	27.990
5:vostorg	25.993	20.455	**13.220**	27.990	0.000