Ways of trying in Russian: clustering behavioral profiles

DAGMAR DIVJAK and STEFAN TH. GRIES*

Abstract

This article proposes a methodology for addressing three long-standing problems of near synonym research. First, we show how the internal structure of a group of near synonyms can be revealed. Second, we deal with the problem of distinguishing the subclusters and the words in those subclusters from each other. Finally, we illustrate how these results identify the semantic properties that should be mentioned in lexicographic entries. We illustrate our methodology with a case study on nine near synonymous Russian verbs that, in combination with an infinitive, express TRY.

Our approach is corpus-linguistic and quantitative: assuming a strong correlation between semantic and distributional properties, we analyze 1,585 occurrences of these verbs taken from the Amsterdam Corpus and the Russian National Corpus, supplemented where necessary with data from the Web. We code each particular instance in terms of 87 variables (a.k.a. ID tags), i. e., morphosyntactic, syntactic and semantic characteristics that form a verb's behavioral profile. The resulting co-occurrence table is evaluated by means of a hierarchical agglomerative cluster analysis and additional quantitative methods. The results show that this behavioral profile approach can be used (i) to elucidate the internal structure of the group of near synonymous verbs and present it as a radial network structured around a prototypical member and (ii) to make explicit the scales of variation along which the near synonymous verbs vary.

Key words: (near) synonymy, behavioral profiles, ID tags, (hierarchical agglomerative) cluster analysis, t-values, z-scores, Russian, verbs of trying

1. Introduction

An intriguiging meaning relation in natural language is that of "(near) sameness of meaning", i. e., (near) synonymy. Synonymy has received

relatively little attention in Western linguistics in recent years. It is said to "waste" the limited lexical resources on one and the same semantic unit, and therefore it should not exist in an ideal one-to-one semiotic system (De Jonge 1993: 521; Taylor 2003: 264). But, even if synonyms name one and the same thing, they name it in different ways; they present different perspectives on a situation. And this provides interesting information on how a particular semantic and related conceptual space is structured. In what follows we will show how a quantitative corpus linguistic approach that is informed by findings from cognitive linguistics provides a solid empirical basis for theoretical modeling.

1.1. Problems of research on near synonymy

Near synonymy is an area in which the theoretical interests of the lexical semanticist and the applied interests of the lexicographer converge. It is, however, also a particularly problematic area, both from a general, lexical-semantic point of view as from a more specific, synonymy-related stance.

Defining any word's meaning or distinguishing between its senses is a rather elusive endeavor; as a consequence, cases of ambiguity and vagueness are difficult to deal with on a principled and objective basis. This situation would be even aggravated when the semantic tests used for distinguishing senses of a single word, i. e., cases of polysemy, were applied to the study of groups that contain two or more semantically similar words, i. e., synonymy. Since no two words ever are exact synonyms, but instead always differ from a syntactic, semantic and/or pragmatic point of view, among other things, scholars assume the existence of a scale of synonymy (Cruse 1986: 267−268; Taylor 2003: 265). Although such a scalar view on synonymy obviates the need for clear-cut decisions, it leaves the analyst with a multitude of possible scalar distinctions to choose from. Some of these problems, which are germane to synonymy research, are summarized briefly in what follows. As examples, we will use tentative verbs (i. e., verbs that express *try*-ing) in Russian.

1.1.1. The delineation problem

The first main problem of synonym research concerning word X, e. g., a Russian tentative verb, is how to decide which near synonyms of X should be mentioned in X's entry and which ones should be left out. A check of three major works dealing with tentative verbs reveals that none of them lists exactly the same verbs.

Consider as a first example, Apresjan et al. (1999: 303−308), who − following the principles of the semantic metalanguage (cf. below for some details) − list only the verbs in (1) as verbs that share the meaning

"make an effort in order to carry out a certain action, while the subject or the speaker does not know whether the effort in question will lead to the necessary result":

(1) *probovat'* ('try'), *pytat'sja* ('try, attempt'), *starat'sja* ('try, endeavor'), *silit'sja* ('try, make efforts')

By contrast, Apresjan et al. (1999: 308) list the verbs in (2) as also exhibiting a significant part of the general semantic structure attributed to the verbs in (1), although there is not enough overlap for them to be considered full-fledged near synonyms of the verbs in (1):

(2) *dobivat'sja* ('get, obtain'), *domogat'sja* ('seek, solicit'), *chotet'* ('want, intend'), *namerevat'sja* ('intend, mean'), *stremit'sja* ('strive, try'), *rvat'sja* ('strain, burst to'), *poryvat'sja* ('try, endeavor'), *bit'sja* ('struggle'), *osilit'* ('manage'), *tščit'sja* ('try, endeavor'), *pyžit'sja* ('go all out'), *norovit'* ('try, strive to, aim at'), *ispytyvat'* ('test'), *probovat'$_2$* ('test')

A second work, Černova (1996: 87), counts nine verbs that fit the 'try, attempt'-stage of the frame 'plan − accomplishment of the plan'. They are listed under (3):

(3) *probovat'* ('try'), *pytat'sja* ('try, attempt'), *starat'sja* ('try, endeavor'), *norovit'* ('try, strive to, aim at'), *silit'sja* ('try, make efforts'), *tščit'sja* ('try, endeavor'), *iskat'* ('look for, seek'), *domogat'sja* ('seek, solicit'), *ne počesat'sja* ('(not) to be itching to')

Černova (1996: 87) states that "a situation of trial and attempt arises when the subject is not convinced that s/he will reach the desired result. This uncertainty is brought about by the presence of external or internal obstacles and by the subject's lack of experience to carry out the action".

Finally, the dictionary of synonyms by Evgen'eva (2001, 2: 323, 496) includes as synonyms of *pytat'sja* ('try, attempt') only *probovat'* ('try'). The six verbs in (4) are paraphrased as meaning 'make an effort in order to obtain or realize something' and are given as synonyms for *starat'sja* ('try, endeavor'):

(4) *stremit'sja* 'strive, try', *pytat'sja* ('try, attempt'), *norovit'* ('try, strive to, aim at'), *silit'sja* ('try, make efforts'), *tščit'sja* ('try, endeavor'), *pyžit'sja* ('go all out')

Comparing the verbs presented in Apresjan et al. (1999) with those listed by Černova (1996) and Evgen'eva (2001) exemplifies a typical

problem within synonymy research: different (groups of) researchers arrive at rather distinct sets of near synonymous verbs. Apresjan et al.'s core group of tentative verbs consists of four verbs, while that of Černova's consists of nine and Evgen'eva distinguishes two groups containing two and six verbs respectively. In other words, Apresjan draws the line for near synonyms, listed in (1), much tighter than the other two researchers. At the same time, he includes reference to groups of "semantically similar verbs", enumerated in (2); the semantic criteria for inclusion in this latter type of group are much looser, as already becomes apparent from the translations that, apart from verbs that express *try*, include verbs like *test*, *struggle* and *manage*.[1] The method to be outlined below will address this problem of delineation in more detail.

1.1.2. The structuring problem

The comparison of the three researchers' groups of tentative verbs does not only show that the groups differ in terms of the number of verbs that are considered as near synonyms − it also reveals differences concerning how the groups of verbs are structured. For example, Apresjan et al. (1999: 303−308) treat *probovat'* and *starat'sja* as belonging to a small core group of tentative verbs whereas Černova (1996: 87) classifies them as belonging to a much larger group of tentative verbs. Evgen'eva (2001: 323, 496), however, keeps *probovat'* separate from nearly all other verbs, which are classed together with *starat'sja*. Conflicting analyses such as these underscore the general need for more objective and thus replicable lexical-semantic analyses. The internal structure of a group of near synonyms is an issue that has hitherto remained largely undiscussed in the literature. With the notable exception of Edmonds and Hirst (2002), many if not most analyses we are aware of tend to treat synonyms in pairs; cf. standard textbook references (cf., e. g., Cruse 1986; Saaed 1997), lexical-semantic studies (cf., e. g., Geeraerts 1985; Mondry and Taylor 1992), corpus-based studies (cf., e. g., Gries 2001, 2003; Kjellmer 2003; Taylor 2002), etc. However, synonym dictionaries and thesauri typically cross-reference individual words, thus revealing that pairs of near synonyms form larger series of semantically similar words and word fields. This raises the question of whether and, if so, how semantically coherent categories are structured internally. We will take up this question below.

1.1.3. The description problem

The third problem is one of comparing potentially synonymous words. A prerequisite for measuring the similarity between words is having a

means to compare them. While traditional semantic analysis has advanced several tests to uncover subtle differences between senses of lexemes and/or between lexemes (cf. Cruse 1986 for insightful discussion) these tests offer relative judgments, e. g., which senses and/or lexemes are *more* or *less* similar in meaning, without specifying the magnitude of the difference and providing an objective motiviation for it. In addition, given that these tests are not designed to produce precise results, they have only low interrater reliability and replicability. What is needed, therefore, is a reliable means to capture differences between different words' meaning.

1.2. Objectives and overview of the present paper

In this paper, we develop and outline a largely objective and verifiable approach to tackle two of the three above-mentioned problems, i. e., structuring and description (for delineation see Divjak 2004, in press).

The paper is organized as follows. Section 2 presents our methodology in quite some detail. More specifically, we will comprehensively discuss the corpus data we investigate (Section 2.1) as well as all the variables making up a behavioral profile (Section 2.2). In addition we will introduce the statistical method of hierarchical agglomerative cluster analysis, which we use for inferring structure from the data (Section 2.3). Section 3 presents the results of cluster analysis applied to the full set of variables. Section 4 illustrates how the results from the cluster analysis and additional statistics derived from the behavioral profiles can be used to develop a radial network of the verbs investigated; in addition, it shows how this approach facilitates identifying subtle semantic differences between near synonymous verbs. Finally, Section 5 summarizes our main results and outlines possibilities for future research.

We use the example of near synonymous tentative verbs in Russian (introduced above) to exemplify our method, but wish to emphasize from the outset that it is applicable to other areas of lexical semantics as well; in fact, the method has already been applied successfully to the analysis of a highly polysemous English verb, *run* (cf. Gries 2006). In other words, this study has a theoretical and methodological focus: it aims at elucidating the structure of the category of tentative verbs and finding the elements that are of interest for describing the prototype usage of each verb − it does not pretend to offer a full-fledged "lexicographical portrait" (Apresjan et al. 1995) of each tentative verb.

In this paper, we approach near synonyms from a quantitative, corpus-linguistic perspective. More specifically and like most other corpus-linguistic approaches to lexicography/semantics, we make use of an assumed correlation between *distributional* similarity on the one hand and

semantic similarity on the other hand.[2] To that end, we have implemented the idea of correlating distributional and semantic similarity to a larger extent than what would be considered standard in corpus linguistics. Our principal method of investigation is to extract every clue possible (which we, following Atkins 1987, refer to as "ID tags") from the corpus sentences in which the verbs under consideration are used so as to infer different facets of their meanings. These ID tags comprise formal characteristics of the finite verb and the clause or the sentence the finite verb occurs in, elements that co-occur with the tentative verb (such as adverbs, particles and connectors) as well as paraphrases (i. e., characterizations) of the semantic properties of the subject and infinitive. Taken together, these ID tags form what we, taking up a term coined by Hanks (1996: 79), refer to as the "behavioral profile" for each verb.

Given the distributional orientation of this study, it is necessary to point out how our research relates to work on (near) synonymy within computational linguistics. However, compared to the vast number of works on what one can consider the computational linguistic approach to polysemy, i. e., word sense disambiguation, there is much less work on (near) synonymy (cf. Edmonds and Hirst 2002: 106 for a similar assessment). In addition, much of the work we are aware of (i) proceeds on the basis of lexical co-occurrence data alone as opposed to the richly annotated data we will introduce below and (ii) has a less theoretical set of objectives (e. g., synonymy identification from text, the automatic generation of thesauri) than the present study. To name but a few, Li and Abe (1998) address the problem of thesaurus construction on the basis of lexical co-occurrences, Turney (2001) compares the performance of different unsupervised learning algorithms for recognizing synonyms using lexical co-occurrence, and Pearce (2001) uses synonyms coded in WordNet as a mere tool for collocation extraction.

Our study differs from previous analyses of synonymy in at least one of the following three ways. First, we aim at making the analytic process as *precise* as possible, which is why our entire data set consists exclusively of manually annotated corpus data; this contrasts with computational linguistic studies in which larger data sets are used but where automatic coding may result in lower degrees of precision and recall. Indicative for this lower degree of precision that comes with automatic tagging is the fact that even in areas largely relying on automatic analysis, manually tagged data sets are still used as gold standards (cf., for example, Evert and Krenn 2001). Second, we aim at making the analytic process as *objective* as possible, which is why our manual annotation is largely, though not fully, restricted to objectively identifiable properties. The properties themselves were chosen as being relevant for shedding light on covert semantic properties on a theoretical basis, a topic we will return to

below. Third, we aim at making the analytic process more *comprehensive* than many previous studies in two ways. On the one hand, the number of properties we have included in our analysis is much larger than the number found in most traditional semantic or lexicographic analyses (cf. below for details). On the other hand, we do not restrict our attention to pairs of semantically similar words but include a larger range of candidate words that share similar constructional properties, as (5) reveals.

2.   Methods

2.1. The data

The group of tentative verbs that is focused on in this case study comprises the nine tentative verbs listed in (5):[3]

(5) *probovat'* ('try'), *pytat'sja* ('try, attempt'), *starat'sja* ('try, endeavor'), *silit'sja* ('try, make efforts'), *norovit'* ('try, strive to, aim at'), *poryvat'sja* ('try, endeavor'), *tščit'sja* ('try, endeavor'), *pyžit'sja* ('go all out'), *tužit'sja* ('make an effort, exert oneself')

The list has been assembled on the basis of the absence of the three constructional properties that typically characterize finite verbs in Russian [V_{FIN} V_{INF}] sequences (Divjak 2004, 2006). Contrary to the majority of verbs that combine with an infinitive, these 9 tentative verbs can neither pull the infinitive into one of their argument structure slots, nor can they subordinate the infinitive event or let the infinitive event be modified with temporal specifications that diverge from the ones added to the finite verb.[4] On the distributional hypothesis, there are systematic meaning differences to be discovered between verbs that display one, two, or all three constructional characteristics and verbs like the tentative verbs treated here that do not have any of these properties (cf. Atkins and Levin 1995: 96). Therefore, verbs that do not lack all three properties are not taken into account.[5]

Our corpus based analysis has been carried out on a collection of 1,585 sentences that contain one of the nine tentative verbs mentioned above. The main source of data is the ten-million-word section (10,750,757) of the Amsterdam Corpus (AC) that contains literary works from different genres, originally written by appr. 75 authors in Russian between 1950 and 2000, complemented with data from the Russian National Corpus (RNC) extracted from literary works from the same period, and where necessary from the Web. The exact numbers of examples used are given in Table 1.[6] Ožegov and Švedova (1999) mark *silit'*sja, *norovit'*, *pyžit'sja* and *tužit'sja* as "spoken language", and label *tščit'sja* "bookish". Evgen'eva (2001) classifies *pyžit'sja* as "common parlance".

Table 1.   *Corpus examples used per tentative verb*

Verb	N (AC/RNC /Web)	Verb	N (AC/RNC /Web)
probovat'	246 / − / −	*poryvat'sja*	31 / 88 / −
pytat'sja	247 / − / −	*tščit'sja*	21 / 30 / 21
starat'sja	248 / − / −	*pyžit'sja*	− / − / 98
silit'sja	57 / 185 / −	*tužit'sja*	− / − / 53
norovit'	112 / 148 / −		

2.2. Annotation parameters

All 1,585 sentences were tagged for in total 87 variables that cover virtually all the clues encountered in the total of sample sentences provided. That is to say, the analyses that follow are based on a set of 137,895 manually coded data points;[7] this size is − to the best of our knowledge − enormous compared to other analyses of near synonyms that are *not* based on automatic annotation. To quote just one example, Taylor's (2003) analysis of *high* and *tall* is based on 638 data points (the nouns modified by 638 occurrences of *high* and *tall* in the LOB).

The behavioral profile approach we advocate has several advantages. First, it incorporates Vanhatalo's (2003) findings from population tests, showing that the choice of near synonyms is most probably affected by a broader cotext than simply two contiguous words or contiguous collocations. Second, the behavioral profile approach forces one to code all the corpus examples for the same set of parameters, which ensures that all the information that enters into the analysis is made explicit and intangible issues do not sneak in through the intuitive backdoor. Finally, we submit that − for fine-grained lexico-semantic analyses of the present kind − the analysis of words occurring in an arbitrarily defined number of words within a window around the word(s) of interest is inferior to a syntactically-defined range such as the clause or the sentence containing the search word(s) or just one particular constructional slot as in collostructional analysis (cf. Stefanowitsch and Gries 2003; Gries, Hampe, and Schönefeld, 2005, for examples and discussion of the latter approach). In the absence of annotated intonation units, sentences or clauses − as the formal counterparts to propositions − can be considered the most natural units of speech production. Moreover, given that we attribute a central status to distributional information of syntactic and semantic nature, we must take the syntactic structures seriously in which the verbs under consideration are used as well as the types of modifiers and the range of collocates these structures harbor.

2.2.1. *Annotation of formal properties*

Coding started from observable formal characteristics of the finite verbs and was gradually extended to include information on other elements of the sentence. In a first coding round, we zoomed in on the elements present in all constructions built on the $[V_{FIN} V_{INF}]$ pattern. Finite verbs do not exist without a specific aspect, mode and tense. Therefore we propose to take these three characteristics and their different values presented in Table 2 as the basis for further analysis.

Table 2. *Verb related information*

Tentative verb			Infinitive
Aspect	Mode	Tense	Aspect
imperfective, perfective	infinitive, indicative, imperative, participle, gerund, conditional	future, present, past, not applicable	imperfective, perfective

Since these verb related characteristics invariably occur together, we have correlated the aspect, mode and tense of the finite verb with the aspect of the infinitive. There are two reasons for doing this. First, the verbs of this study owe their tentative character to the particular construction they appear in, i. e., the combination with an infinitive. Second, aspect is a rich source of information for the semantics of a verb. Tagging for aspectual preferences results in a degree of semantic precision that is much harder to achieve for languages that do not express aspect morphologically. The current state of research on aspect, however, does not make it possible to state explicitly all motivating components for a particular aspect and how they relate to the semantics of individual verbs (Janda 2004). Apart from aspect, mode and tense, we have added those elements that are strictly necessary to form a full-fledged sentence, i. e., the case marked on the subject slot and information on the type of clause the $[V_{FIN} V_{INF}]$ sequence is used in. These coding possibilities, which cover all clause, sentence and subject types encountered in our sample, are presented in Table 3.

Taken together, the structural data on clause type and related form of the subject, as well as details on the aspect, mode and tense of the verbs in the $[V_{FIN} V_{INF}]$ sequence, form the skeleton of the sentence. From here, one can fill up constructional slots with lexical elements. The idea of constructional slots is of special importance for the analysis of semantically related and near synonymous verbs. Due to the high similarity and low contrastiveness that characterizes near synonyms they are usually defined as highly intersubstitutable (Cruse 1986: 268). As basic pre-

Table 3.   *Clause and subject related information*

clause structure	main clause, subordinate clause
sentence type	declarative. interrogative, imperative, exclamative
subject structure	nominative to the tentative verb, nominative to the preceding verb, accusative to the preceding verb, dative to the preceding "personal" verb, dative to the preceding "impersonal" verb, dative to the tentative verb, the subject is the infinitive tentative verb, the infinitive tentative verb modifies a noun

requisite for intersubstitutability similarity of context is often named (Cruse ibid.). In this paper, similarity of context has been supplemented with similarity of construction: the more slots in a construction, here the [V_{FIN} V_{INF}] pattern, coincide in building-up and the more frequently these slots are used in the same way, the higher the chances of intersubstitutability between two or more of these near synonyms. It is surprising that this factor has been largely neglected in previous studies on near synonymy. To the best of our knowledge, only Arppe and Järvikivi (2002) discuss the reality of selectional differences based on morphosyntactic features in relation to near synonyms. In studies on polysemy, on the contrary, the exact form of a verb and its context have recently been introduced (cf. Divjak 2006; Gries 2006).

From the second coding round onward, the tagging is inspired by the meticulous "portrayal" methodology, developed within the Moscow School of Semantics (Apresjan 1995 et al.). The emphasis now lies on a thorough analysis of the variation in combinatorial preferences of the nine tentative verbs. Combinatorial patterns are prevalent in corpus linguistics and are typically interpreted as reflecting semantic and/or pragmatic characteristics. However, as explained above, we do not restrict the window of analysis to an arbitrarily chosen number of preceding or following words, but focus on every element encounted in the simple sentence or clause the tentative verbs occur in.

In the second round, the adverbs, particles and connectors that are used in the corpus sample are at the center of attention. Detecting adverbs, particles and connectors does not require semantic intuitions, but is semantically informative (cf. Section 4.1.2). Going through all examples reveals that the set of adverbs, particles and connectors actually used is quite limited. In other words, verbs combine with a whole range of adverbs, particles or connectors, but not all verbs prefer identical (sets of) adverbs.[8] As Table 4 shows, the adverbs encountered in constructions with a tentative verb tend to stress the attempt itself. More precisely, these adverbs provide information on how long, how often, how intensely the subject tried, and whether his/her attempts were successful.

Table 4. *Adverbs, particles and connectors*

adverbial specification	duration (*dolgo* 'long', *dolgoe vremja* 'a long time' …), durative repetition (*vsë* 'all (the time)', *vsë vremja* 'all the time' …), repetition (… *raz* '[…] times'), intensity (*očen* 'very', *izo vsech sil* 'with all one's might' …), vainness/futility (*zrja, naprasno, tščetno* 'in vain' …), intensity and vainness (*kak ni/ne* … 'however')
particles	exhortation (*davaj* … 'let's, come on'), permission (*pust'* … 'let'), restriction (*tol'ko* … 'only, just'), permission and restriction (*pust' tol'ko* … 'let … only'), intensification (*daže* … 'even'), untimely halt (*bylo*)
negation	to the tentative verb, to the infinitive
connectors	external opposition (*no, a, i ne*), internal opposition (*no, a, i ne*), introducing a *čtoby* 'in order to' clause, in a *čtoby* 'in order to' clause

2.2.2. *Annotation of semantic properties*

The third type of information that has been coded is the one most semantic in nature. It contains semantic paraphrases for the subject and the infinitive, typical candidates for a traditional collocation analysis in languages with fixed word order such as English.

Within the scope of this study, we systematically classified the nominative subject paradigms along a combination of lines, i.e., the opposition animate vs. non-animate including insects and the distinction between addressable, i.e., human, and non-addressable or animal animate subjects. We introduced some additional distinctions for non-animates, i.e., concrete vs. abstract and further specifications have been made on the basis of the kinds of subjects used in the data sample, i.e., man-made and non-man-made concrete things, the latter being most often phenomena of nature (e.g., *the sun, the earth*) or body parts, as well as abstract concepts (e.g., *an idea, an insight*) and groups or organizations.

For the infinitives, we adopted a labeling system that is inspired by the "semantic primitives of human behavior" set forth in Apresjan's (1995a) linguistic naïve world view. The eight "basic systems of a human being" that Apresjan (1995a: 355−356) distinguishes are comparable to "basic domains" (Langacker 1987: Chapter 4) or the semantic primitives underlying the Natural Semantic Metalanguage by Wierzbicka (1996), i.e., domains that are not characterized in terms of other more fundamental domains. These semantic primitives of human behavior are generalizations over the paraphrasing semantic labels for classifying infinitives we use in our research. The details are presented in Table 5.

Verbs that are related to the primitive for "physical actions" are extremely frequent, which might make this coverterm too vague, thus obscuring meaningful distinctions. Therefore "physical actions" are subdi-

Table 5.   *Apresjan's eight primitives of human behavior related to corpus tags*

Apresjan's domains	Apresjan's primitives	Coding labels used
physical actions	*delat'* 'do'	physical, physical other, physical exchange/transfer, physical motion, physical motion other, physical figurative, physical figurative other, physical exchange/transfer figurative, physical motion figurative, physica motion other figurative
physical perception	*vosprinimat'* 'perceive'	perceptual, perceptual active
speech	*govorit'* 'say'	communication/interaction
intellectual activity	*znat'* 'know', *ščitat'* 'consider, think'	mental
emotions	*čuvstvovat'* 'feel'	emotional
physiological condition	*oščuščat'* 'feel, sense'	(absent from our sample)
physiological reactions	[no primitive available]	(absent from our sample)
wishes, desires	*xotet'* 'want'	(absent from our sample)

vided on the basis of their argument structure schemes. The label "physical" is reserved for verbs that do not have an accusative slot, i. e., actions that only involve the subject, e. g., *get dressed*, *sleep*. "Physical other" covers verbs that do have an accusative slot that is affected, as in *beat someone*, or even ceases to exist as in *kill someone*. A second group of verbs that take objects are "physical exchange/transfer" verbs. These verbs have an object slot that is transferred but remains unaffected by the action, e. g., *sell something*, and easily introduce a third participant in the form of e. g., a dative *to someone* (also cf. Fisher et al. 1991: 379, who term this type of three-place predicates "transfer"). This change reminds of activities that involve motion: motion of the subject itself ("physical motion"), and motion of an object ("physical motion other"), parallel to the affected self and the affected other. For all five basic groups, "physical", "physical other", "physical exchange/transfer", "physical motion", and "physical motion other", there are figurative categories that contain all the verbs that instantiate these schemas, though in another domain than the physical one; in other words, these are usually metaphorical uses (in, say, Lakoff's 1987 sense of the term). At the same time, the figurative uses have not developed a relation to another

semantic primitive either. Physiological reactions, physiological conditions and wishes or desires are not represented in the tags: there were no examples that contained verbs proportional to given primitives.[9]

Of all other primitives, only the perceptuals were further divided into active, conscious, directed perception as in *look* and passive, accidental perception like *see*, as Apresjan (1995a: 357) proposes. The latter feature of directed versus accidental action is closely related to an additional important parameter, i. e., the degree of (objective) controllability of the infinitive actions, introduced in Apresjan et al. (1999: 304−305). The tests we used in tagging for this variable are inspired by the Vendlerian (1967) stativity tests. Controllable verbs such as *copy* pass the imperative test that can be seen to include the "persuade to Y" and "force to Y" test and combine with the adverb-tests "to Y deliberately" or "to Y carefully". Weakly controllable verbs such as *find* pass the imperative tests but do not pass the adverb tests, e. g., you can *order someone to find something* but you cannot *find something carefully* or *deliberately*. Verbs like *forget* that do not pass either test are tagged "non-controllable".

Many other properties could be tagged for. However, either some of these other tags are already included indirectly, or they do not apply in this particular case of verbs, or do not lend themselves well to operational definitions. We will discuss each of these options in turn.

It has often been pointed out in the literature (for an overview see Janda 2004) that aspect in Russian relates to Vendlerian verb classes. Therefore, Vendlerian-type event characteristics are not included as separate tags, as this would double the impact of aspectual parameters. Vendlerian verb classes are at the same time an example of non-applicable tags: although stativity of the infinitive seems to be one of the factors for distinguishing, e. g., between *stat'* ('begin, start') and *načat'* ('begin, start'), it does not yield interesting differentiating results for these nine tentative verbs: none of the tentative verbs combines with statives. Since the goal of this case study is to point out which distributional properties can be used to form subcategories within a group of verbs that are constructionally and semantically highly similar, we have directed attention to characteristics that differentiate between the nine verbs, as is done within the Moscow Semantic School (Apresjan et al. 1995, 1999, 2000; Rachilina 2000). Finally, some of the distinctive features Apresjan et al. (1999: 303−304) enumerate, e. g., the "difficulty of the action", the "objective need for a certain amount of effort" and the "subject's motivation for undertaking the action" lack operational definitions. This is also the case for the "visibility" of the effort. These features, except for an occasional mentioning of visibility, are not explicitly expressed in the sentences. Neither is it clear which tests could be

introduced to measure something like difficulty. Fisher et al. (1991: 375 ff.) point out that the more refined the semantic information, the less clear the formal patterns are. Therefore, such characteristics were not included in the analysis. Finally, tags that relate to attitude or emotional value, register and style remain outside the scope of this analysis (cf. Edmonds and Hirst 2002, who likewise model these properties outside the denotational level).

In sum, in operationalizing the tenet that "the semantics of each verb in the language are determined by the totality of its complementation patterns" (Hanks 1996: 75) we draw mainly on achievements of the Moscow Semantic School (Apresjan et al. 1995, 1999, 2000): not only did they develop this methodology to a larger extent, but also implemented it in the volumes of the *New Dictionary of Synonyms*. Contrary to this school, we examine the distribution patterns exhaustively and take into account the relative frequency they occur at. The results from the 1,585 sentences are for each parameter summarized per verb, yielding a spreadsheet with 87 values (expressed in percentages) for each of the nine verbs. In the next section, we will introduce the method used for analyzing the data and the main results it produced.

2.3. The method: hierarchical agglomerative clustering

The large volumes of data, and the assumption that all elements in the sentence have a valuable contribution to make in providing clues for the subcategorization and related lexical-semantic description of the nine tentative verbs, make a compelling case for a quantitative approach. It is unlikely indeed that a human analyst could keep track of all 87 variables when computing the similarities and differences that indicate basic tendencies. Therefore, hierarchical agglomerative clustering is used to organize the data, determine degrees of similarity between the nine verbs, and obtain smaller clusters that facilitate comparing the elements in the structures to each other (cf. Divjak 2006; Gries 2006).[10]

Hierarchical agglomerative cluster (HAC) analysis is a family of methods that aims at identifying and representing (dis)similarity relations between different items (a general comprehensive discussion of clustering can be found in Kaufman and Rousseeuw 1990).[11] Usually, clustering is performed on the basis of variables that characterize the items or on the basis of a (dis)similarity matrix of the items. The latter can be obtained by tagging corpus extractions for the variables listed above or from similarity judgments or sorting tasks. In our case, the nine verbs correspond to the items while the 87 ID tags expressed as relative frequencies function as variables.

A HAC analysis of the kind used here begins by considering each of the n items as one-item clusters and proceeds to amalgamating those clusters which exhibit the highest intra-cluster similarity and the lowest inter-cluster similarity successively until all clusters have been amalgamated into a single cluster containing all items. The structure yielded by this amalgamation process is typically represented by means of a so-called dendrogram, i.e., a tree diagram representing the similarities among clusters. In addition, a variety of statistical measures can be outputted that help (i) to determine the number of clusters one should assume as well as (ii) to identify which of the variables drive the clustering.

There are two issues that merit brief discussion. First, while the outcome of a cluster analysis can be submitted to further statistical testing (examples include identifying properties that discriminate between clusters by means of t-values, identifying properties that discriminate between items within a cluster by means of z-scores, and a variety of other measures; cf. below for an application of the techniques mentioned) cluster analysis is an exploratory, hypothesis-generating technique and strict significance testing is typically not applied. Second, the precise outcome of a cluster analysis is contingent on at least two important settings, namely (i) the measure of (dis)similarity one chooses to assess intra- and inter-cluster similarity and (ii) the amalgamation strategy one selects to determine how the clusters should be joined on the basis of their (dis)similarity. There is no uniformly accepted combination of parameters that guarantees an optimal clustering solution, however, which is why we have adopted parameters that have been identified as yielding useful results in previous work. Our cluster analysis has been performed using the Canberra metric as a measure of (dis)similarity because it handles the comparatively large number of zero occurrences of particular features best.[12] Ward's rule, based on minimizing squared sums of error, was selected as the amalgamation strategy since it is a frequently recommended strategy that yields small clusters (Ward 1963).

In the next section we will discuss the results of applying HAC analysis and show that the degrees of similarity by and large converge with native speaker intuitions; unless mentioned otherwise, all computations were carried out using R 2.2 (R Development Core Team 2005).

3. Overall results

Taking the percentages obtained for all 87 parameters for all nine verbs on the basis of all the 1,585 sentences in the data sample through the HAC analysis yields the dendrogram in Figure 1.
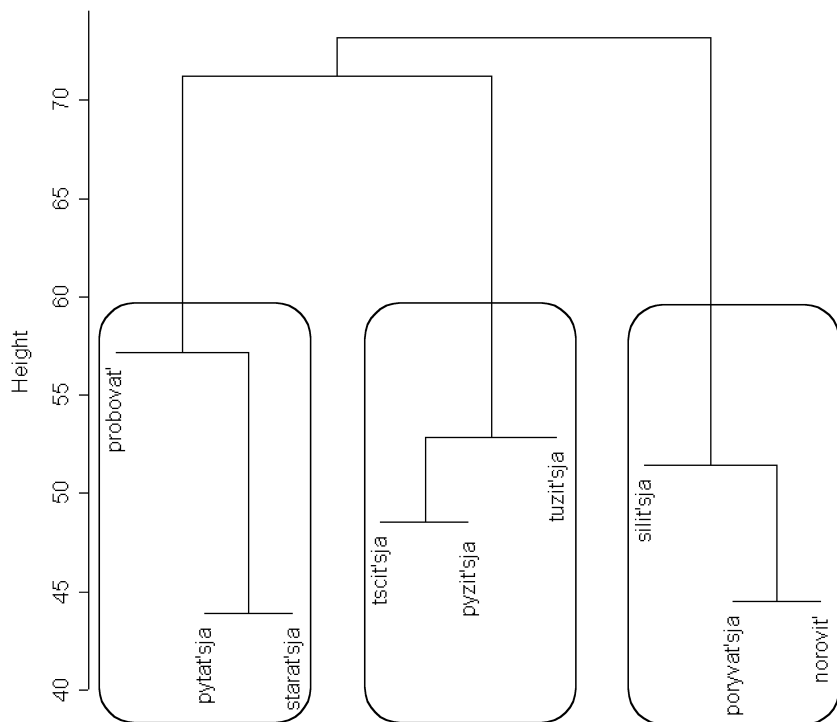
Figure 1.   *Dendrogram for the tentative verbs*

The dendrogram provides two types of information that can be read off the tree plot from bottom to top. On the one hand, the tree plot shows what is similar and what is different: items that are clustered or amalgamated early are similar, and items that are amalgamated late are rather dissimilar (when compared to verbs amalgamated earlier, that is: even the tentative verbs that are amalgamated late in this analysis are still very similar to the tentative verbs amalgamated earlier [cf. below for evidence] as compared to, say, semantically unrelated verbs such as *čitat'* ['read'] or *stroit'* ['build']). On the other hand, the tree plot gives an indication of how independent the clusters are: the larger the distance between different points of amalgamation, the more autonomous the earlier verb/cluster is from the verb/cluster with which it is merged later − for example, it is obvious that *pytat'sja* and *starat'sja* are much more similar to each other than, say, *probovat'* and *norovit'*. Typically, the first most pressing question is how many clusters one should reasonably assume. In the present case, this question can be answered straight-

fowardly: There are three clusters, each of which contains three verbs. The clusters are framed in Figure 1.

In evaluative terms, we could say that − although we do not rely on introspection and potentially volatile intuitions − the results of the corpus-based cluster analysis yield a dendrogram that suggests delineating groups of near synonyms that are similar to Apresjan et al.'s (1999) classification. However, there are also some differences that will be discussed in what follows. Apresjan et al. (1999: 303−308) treat *probovat'*, *pytat'sja*, *starat'sja* and *silit'sja* as near synonyms that cluster around *pytat'sja*. In the present cluster analysis, the first three verbs are indeed grouped together, and *pytat'sja* plays a central role since it is the first verb to be amalgamated at all (with *starat'sja*), but *silit'sja* is not grouped together with the former three verbs. Another difference is that Apresjan relates *probovat'* most closely to *pytat'sja*, whereas the tree diagram pictures *pytat'sja* and *starat'sja* as most closely related. We will discuss this divergence in more detail in Section 4.1.2.

Secondly, the earlier verbs or clusters are amalgamated in a dendrogram, the more similar they are, whereas the later the verbs or clusters are linked to each other, the less synonymous they are. Consequently, e. g., *pytat'sja* and *starat'sja* are closer near synonyms, whereas *probovat'* and *norovit'* are rather different, given that the former are merged immediately while the latter are part of clusters that are only merged near the end of the amalgamation. This is precisely what Apresjan et al. (1999: 303−308) suggest in their analysis. Likewise, in the "analogi" section that contains words that are similar but not similar enough to be considered near synonyms, Apresjan et al. (1999: 308) mention *tščit'sja* and *pyžit'sja* as closely related, which is exactly what the dendrogram of the behavioral profiles indicates.

In the next section, we will show how the results of a cluster analysis performed on the basis of behavioral profiles contribute to a better understanding of the structure of lexical categories and can be incorporated in lexical semantic analyses.

4.   Applications of corpus-based behavioral profiles

4.1. Introduction

A cognitive approach to language and more in particular the incorporation of knowledge about human categorization mechanisms into linguistics provides interesting perspectives for a unifying interpretation of the data. On this approach, (linguistic) categories may exhibit prototype effects and instantiate radial networks of related expressions with semantically motivated connections (Lakoff 1987: Chapter 6). Following Gries

(2006) we present a corpus-based semantic network for the nine verbs of trying. More specifically, in this section we develop a radial network representation that is explicitly modeled on the cluster-analytic dendrogram and support the analysis based on the overall dendrogram with complementary statistics computed on the basis of our data. Our representation is non-committal with respect to issues of how the knowledge or patterns of usage are actually represented in the brain, however.

We use two kinds of additional diagnostics, each of which investigates different aspects of the data. First, we look at similarities and differences *between* clusters. To that end, we computed *t*-values to determine which variables are most strongly represented (in the case of high positive *t*-values) and which variables are most strongly underrepresented (in the case of high negative *t*-values) in a particular cluster (cf. Backhaus et al. 1996: 310−312). Given the large number of results yielded by this procedure, we restrict our attention to the most revealing scores, i. e., scores which highlight the differences between clusters. More specifically, to characterize the most important variables for one cluster and to identify the variables that discriminate best between clusters, we looked at (i) the variables having positive *t*-values for one cluster and negative *t*-values for the other two clusters and (ii) the variables having negative *t*-values for one cluster and positive *t*-values for the other two clusters. These values provide us with the scales of variation along which the three clusters differ most; in the interpretation we focus on the ranking of the *t*-values.

Second, we look at similarities and differences *within* clusters. To that end, we computed z-scores for each variable and verb in each cluster; this way we determine which variables are most strongly associated with one verb as opposed to the others (in the case of high *z*-scores) and which variables are least associated with one verb as opposed to the others (in the case of low *z*-scores). Again, in the interest of space we only focus on the strongest indicators, concentrating on the ranking of the scores rather than focusing on their absolute values.[13]

In the next section, we will show how clustering behavioral profiles and evaluating clusters and verbs in terms of *t*-values and *z*-scores assist in deciding on the prototypicality of one lexeme as well as in answering the question of where to connect a lexeme to another in the network. In addition, we show how ID tags provide us with scales of variation for describing and distinguishing near synonyms in a fine-grained lexical semantic analysis.

4.2. *Radial network representations and scales of variation*

On the basis of the HAC analysis performed on the totality of the behavioral profiles we propose a radial network for near synonymous tentative
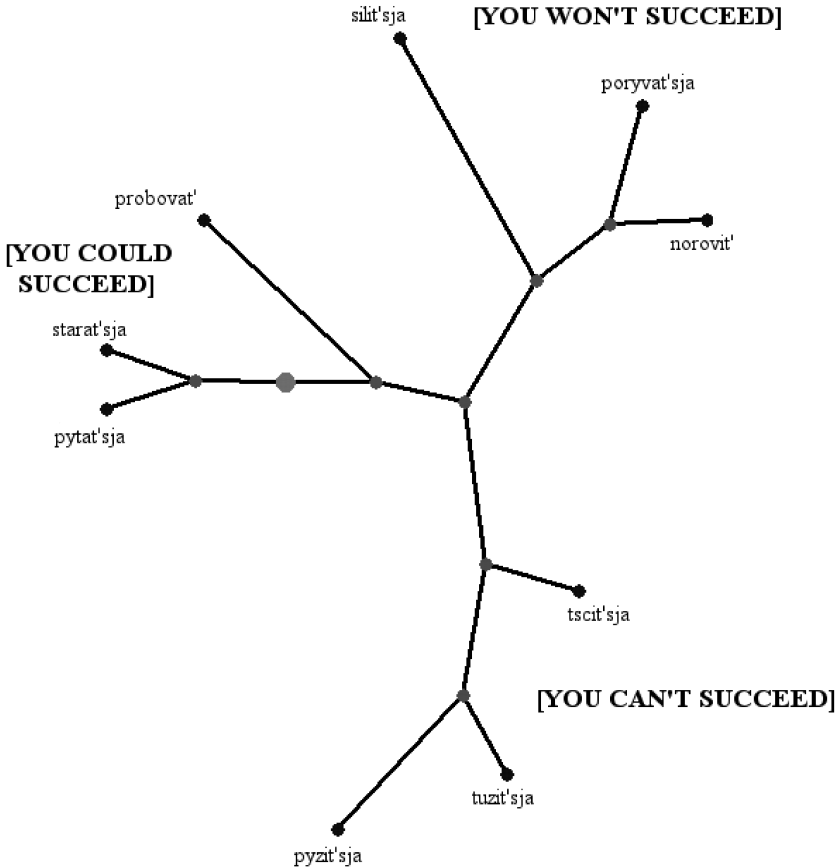
silit'sja

[YOU WON'T SUCCEED]

poryvat'sja

probovat'

**[YOU COULD
SUCCEED]**

starat'sja

norovit'

pytat'sja

tscit'sja

[YOU CAN'T SUCCEED]

tuzit'sja

pyzit'sja

Figure 2.   *A radial cluster tree for tentative verbs*

verbs that consists of the three clusters indicated in Figure 1. Although the HAC dendrogram presented in Figure 1 can be manually transformed into a radial network representation we opted for backing up our results by analyzing the distance matrix resulting from the behavioral profiles using the Fitch program from the PHYLIP package (Felsenstein 2005). Fitch takes as input a distance matrix and attempts to generate a tree in which the distances between elements fit the distances in the distance matrix best. The resulting solution can then be represented as a radial network tree shown in Figure 2.

We will use the HAC dendrogram from Figure 1 together with the radial cluster tree in Figure 2 as the starting point of an analysis[14] in

which, in the parlance of cognitive linguistics, the nine verbs form a radial network of overlapping, yet at the same time distinct, lexical and conceptual systems. The three different clusters the nine lexemes form can be interpreted as representing different Idealized Cognitive Models, i. e., gestalt-like cognitive structures that provide the background against which experience (linguistic or otherwise) is categorized and interpreted (Lakoff 1987: Chapter 4; Hirst 1995) and that at the same time abstract away from scalar subconceptual intercluster differences (cf. Edmonds and Hirst 2002: 115); tentative labels for these ICMs, i. e., YOU COULD SUCCEED, YOU CAN'T SUCCEED and YOU WON'T SUCCEED, have been inserted between square brackets in Figure 2.[15] These definitions are abstractions over the dimensions that receive the highest and lowest *t*-values for each cluster. If we pull together the dimensions with the most revealing *t*-values for the YOU COULD SUCCEED cluster and incorporate them into one scenario, the characterization that emerges for the three most frequently found verbs, *pytat'sja*, *starat'sja* and *probovat'*, is the following: a human (rather than an animal or an insect) is exhorted to undertake an attempt to move himself or others (rather than to undertake mental activities); often, these activities are negated. With verbs from the YOU CAN'T SUCCEED cluster that contains *tščit'sja*, *pyžit'sja* and *tužit'sja*, an inanimate subject (concrete or abstract) attempts very intensely but in vain to perform what typically are metaphorical extensions of physical actions. In the YOU WON'T SUCCEED cluster with *silit'sja*, *poryvat'sja* and *norovit'*, an inanimate subject undertook repeated non-intense attempts to exercise physical motion; the actions are often uncontrollable and fail because of internal or external reasons.

It is important to mention, though, that these ICMs and their interpretations are stylized characterizations and only serve expository purposes. They comprise those dimensions most relevant for a cluster, yet do not claim that there is any one example that actually incorporates all these criteria simultaneously. This characterization is, in other words, a very specific core scenario, and actual usage events may differ from it along one or more parameters. We will now discuss the properties of each of these clusters and the verbs in the clusters in more detail.

4.2.1. *[YOU COULD SUCCEED]*

The first cluster consists of three verbs [[*pytat'sja* and *starat'sja*] and *probovat'*]. All verbs in this cluster are more easily used in the main clause ($t = 0.821$) than verbs from other clusters. More precisely, they are more easily used in the infinitive, in which case the animate subject belongs to the preceding finite verb (*t*-values range from 0.492 to 1.008). Although all three verbs exist in the imperfective and perfective aspect and occur

in both aspects, variables that include reference to the perfective aspect (i. e., refer to past and future events) are three times more frequent in the top 25 *t*-values that are positive for this cluster and negative for other clusters (*t*-values range from 0.667 to 1.201). In addition, the infinitive that follows the tentative verb is often negated ($t = 0.702$) and expresses physical activities ($t = 0.599$), events that are figurative extensions of motion events ($t = 0.465$) or involve putting an "other" to motion ($t = 0.4$). Finally, optional collocates express that the subject got permission to carry out the infinitive action (using *pust'*, $t = 1.008$), that the attempt was untimely brought to a halt (with *bylo*, $t = 0.982$), that the subject was exhorted to undertake an attempt ($t = 0.832$) and that the intensity with which the attempt was carried out was reduced ($t = 0.667$).

In a behavioral profile analysis of the polysemous verb *run*, Gries (2006, Section 4.1) showed that the verb senses that are amalgamated first are prime candidates for the central senses because they are most frequent both as verbs and zero-derived nouns, they are historically and developmentally primary, and most versatile in terms of the number of ID tags they take. Many of these arguments also apply to our verbs, yet the evidence they provide is ambiguous. First, *pytat'sja* and *starat'sja* are the verbs which are most similar to each other (they have the smallest Canberra distance and are, thus, also amalgamated first in Figure 1) and, relatedly, the cluster they form is at the bottom of the whole hierarchical dendrogram so that all other verbs are ultimately added to them. Secondly, the two verbs are the most frequent of our nine verbs (both in corpora and on the Web) and there is a fairly clear, though by no means absolute, positive correlation between frequency (translating into entrenchment) and prototypicality (cf. Geeraerts 1988: 222; Winters 1990). On the other hand, *probovat'* has the largest number of different ID tags (57 versus the 52 of *pytat'sja* and *starat'sja*) and is thus the prime candidate for being considered the unmarked member. In addition, data from language acquisition also speak in favor of *probovat'* as the central verb. From the Stoll Corpus of Russian first language acquisition we retrieved all instances of the three verbs (and their aspectual counterparts) followed by an infinitive to determine both their frequencies in child speech as well as the time of their first occurrence. While we did not undertake any more precise analysis of the path of acquisition of these verbs, the most obvious finding were that (i) *probovat'* is by far the most frequent verb across all four children, occurring approximately 50% and 72% more often than *starat'sja* and *pytat'sja* respectively, and (ii) *probovat'* and *pytat'sja* are the first of the three verbs to be used by the four children. At the moment we cannot offer an explanation as to why the findings from child and adult speech seem to reflect different central members.

After having looked at how the first cluster differs from the others, let us now turn to the verbs in that first cluster, and we begin with [*pytat'sja* and *starat'sja*] that are clustered first. Both *pytat'sja* and *starat'sja* express that a subject is "making an effort in order to obtain or realize something" (Evgen'eva 2001 2: 496). Although both verbs resemble each other to a large extent, a close inspection of their distributional properties gives us some idea as to how these verbs differ in conveying the meaning TRY. *Pytat'sja* is more strongly attracted to occurring in the past tense (41.3%, $z = 1.122$ for imperfective indicative past, 1.2% with $z = 1.092$ for imperfective participle past and 0.4% with $z = 1.155$ for perfective participle past, all with perfective infinitives) whereas *starat'sja* is relatively more often found in the present tense (7.3% and 6.6% respectively with $z = 1.153$ for both imperfective indicative and gerund present with imperfective infinitives). *Pytat'sja* is not particularly attracted to weakly controllable actions (8.1%, $z = -1.097$) whereas *starat'sja* avoids controllable actions (84.7%, $z = -1.049$). *Starat'sja* combines, among other things, with passive perception verbs (2%, $z = 1.134$) and figurative extensions of the pattern that is used to put an "other" to motion (4% or $z = 1.149$), whereas *pytat'sja* goes well with mental activities (14.8%, $z = 1.139$) and figurative extensions of motion events (10.1%, $z = 1.155$). *Starat'sja* is frequently found with a negated infinitive (20.2%, $z = 1.151$), thus indicating that the subject is avoiding an event that might take place. Easiest to interpret is the verbs' preference for different adverbs. Looking at the adverbs, *starat'sja* is most strongly characterized by adverbs that express repetitive duration (*vsë vremja*, 0.8%, $z = 1.155$), reduced intensity (0.4%, $z = 1.155$), and intensity (5.2% with $z = 1.101$), whereas *pytat'sja* prefers repetition (4.5%, $z = 1.111$).[16] More specifically, if one has already applied *pytat'sja* without success, a possible way to achieve the desired result despite the initial failure is by using what is encoded in *starat'sja* (cf. (6)):

(6) Он убрал Мазера и Леоновича, **постарается то же проделать с Казаковым** *(уже пытался)*, и весьма возможно, с Соя-Серко. [Ф. Незнанский, Ярмарка в Сокольниках]
'He took away Mazer and Leonovič, is trying (hard) [*starat'sja*] to do the same with Kazakov (he has already tried [*pytat'sja*]), and it is very likely, with Soja-Serko.'

Added to [*pytat'sja* and *starat'sja*] is the node [*probovat'*]. This verb occurs preferably in a main clause (91.9% with $z = 1.127$), and is not typically found in declarative clauses (87.4%, $z = -1.148$). Tags that refer to the perfective aspect receive the highest z-scores for [*probovat'*], ranging from 1.003 to 1.155. Although all three verbs in this cluster have

a perfective counterpart formed by means of the delimitative prefix *po-*, i. e., *po/probovat'*, *po/pytat'sja*, and *po/starat'sja*, *po/probovat'* significantly prefers the perfective aspect in 74.8 % of all examples while *pytat'sja* and *starat'sja*, by contrast, significantly prefer the imperfective aspect, i. e., in 79.6 % and 83 % of all cases respectively ($\chi^2 = 222.72$; $df = 2$; $p < 0.001$, Cramer's $V = 0.548$).[17] Related to the more frequent use of perfective forms is the possibility of locating the attempt in the future (13.8 % for *probovat'* with $z = 1.003$ for indicatives in combination with imperfective infinitives and $z = 1.044$ for combinations with perfective infinitives), as well as a considerable relative dispreference for the present tense (3.7 % for *probovat'* with $z = -0.636$ for indicative present with imperfective infinitives and $z = -1.154$ for indicative present with perfective infinitives, $z = -0.632$ for the gerund with imperfective infinitive and $z = -1.107$ for the gerund with perfective infinitive). Finally, *probovat'* is the only verb that is often found in the imperative mode (25.6 % for perfective imperatives, 1.2 % with imperfective infinitives with $z = 1.092$ and 60 or 24.4 % with perfective infinitives, $z = 1.134$). In interpretative terms, the node [*probovat'*] uses the perfective to present each try as a completed entity. This allows the subject to change method or strategy between attempts, which might be what makes this verb resemble experiments (cf. Wierzbicka 1988: 309; Apresjan et al. 1999: 304). An experimental attempt is also demanded more easily from another person than attempts that require long and/or intense effort, hence the higher frequency of the imperative and attraction of exhortative particles (1.6 % with $z = 1.121$). Likewise, restrictive and permissive restrictive particle combinations receive high positive z-scores (1.2 % and 0.8 %, $z = 1.155$ for both) for *probovat'*, indicating that the subject should try and find out for himself. Moreover, an experimental attempt can likewise not be carried out at all (4.9 % contain a negated tentative verb, with $z = 1.121$). Failure can be attributed to internal and external factors alike (4.9 %, $z = 1.155$ and 11 %, $z = 1.151$). In all, *probovat'* seems to be less intensive than *pytat'sja* (and *starat'sja*) indeed, as example (7) shows:

(7) **Бим** *уже* **пробовал на нее наступить**, но пока еще так, *немножко — только пробовал*. [Г. Трупольский. Белый Бим черное ухо]
    'Bim had already tried [*probovat'*] to step on her, but just like that, a little bit, he had only tried [*probovat'*].'

The six verbs that radiate away from the central node [[*pytat'sja* and *starat'sja*] and *probovat'*] differ in the way they convey and elaborate the central concept of "trying to succeed in carrying out an action". Both remaining clusters contain verbs that only exist in the imperfective aspect.

4.2.2. [*You Can't Succeed*]

In the middle, there is a cluster that unites the imperfective verbs [*tščit'sja* and *pyžit'sja* and *tužit'sja*], which differ stylistically: *tščit'sja* is bookish whereas the two remaining verbs belong to "colloquial Russian" (Ožegov and Švedova 1999). All three verbs lack a perfective counterpart and prefer the present tense more than verbs in other clusters ($t = 1.047$ for perfective infinitives and $t = 0.711$ for imperfective infinitives). Among the most strongly represented variables we encounter the verbs' compatibility with inanimate subjects, both concrete and abstract (t ranges from 1.108 to 1.276), as well as with groups or institutions ($t = 1.297$). Actions are physical ($t = 0.176$) that may affect an "other" ($t = 0.352$) or are metaphorical extensions of physical actions ($t = 0.999$) or physical "other" actions ($t = 0.175$). Focus is on the vainness ($t = 0.962$ for vainness combined with intensity) of the durative effort ($t = 0.750$ for duration adverbs).

Apart from similarities, there are within-cluster differences between these verbs that allow them to convey subtly different views on a situation. Let us start with the most similar verbs, *tščit'sja* and *pyžit'sja*, that are clustered first in the hierarchical solution. *Tščit'sja* is more frequently than the other verbs used in a subclause (45.8%, $z = 1.065$) instead of in a main clause (54.2%, $z = -1.065$). With *tščit'sja*, the subject puts in effort to overcome the hindrances: *tščit'sja* shares its root with *tščatel'-nost'* ('thoroughness, care') (Ch. Townsend, p.c.) that can be needed to carry out an action. Non-man-made things are relatively more often used as subjects (4.2%, with $z = 1.153$) of weakly controllable actions (22.2%, with $z = 1.154$, controllable actions score negatively, $z = -1.146$). Actions that discriminate this verb well from others encode physical motion that includes an "other" (2.8 with $z = 1.076$), figurative extensions of physical actions that may include an "other" (9.7% and 22.2% respectively, $z = 1.155$ for both categories), as well as active and passive perception (1.4% with $z = 1.155$ for both). These actions are attempted in vain (4.2%, with $z = 1.153$) and during some period of time (*vsë vremja* occurs in 2.8% of all examples, $z = 1.155$). Reasons for failure are due to subject-external opposition (1.4%, $z = 1.151$). In other words, the infinitive actions *tščit'sja* is directed at are relatively more often specified as hopeless ventures (cf. (8)) and thus doomed to fail:

(8) *Напрасно* **тщился** Яконов доказать свою загруженность, не-
возможность совмещения. [Александр Солженицын. С круге
первом]
'Jakonov tried in vain to show the extent of his workload, the im-
possibility of making an adjustment.'

The polysemous network around *pyžit'sja* contains a 'being puffed up'-component. Yet, apart from human beings and animals, insects and bodyparts likewise function as subjects (1%, with $z = 1.155$). Interestingly with respect to the presence of a puffed up component is the fact that *pyžit'sja* combines well with infinitives that express communication/interaction (23.5%, with $z = 1.153$). In addition, this verb is also more often found negated (5.1%, with $z = 1.641$) and reasons for failure are subject-external (7.1%, $z = 1.151$). Different from the other verbs in the cluster, *pyžit'sja* expresses a durative attempt (10.2% examples with *dolgo* 'long', $z = 1.075$ and 2% with *vsë vremja* 'all the time', $z = 1.155$). Restrictions of the type *tol'ko* 'only, just' are found (1%, $z = 1.155$):

(9) Но смысл-то какой в том, чтоб **пыжиться стать тем**, *кем вы по крови и по воспитанию быть никогда не сможете?*
     [http://www.litera.ru/slova/chuprina/roman4.html/]
     'But is there any sense in this, trying (going all out) to become someone *who you can never be according to your blood and upbringing?*'

Tužit'sja (cf. (10)) is a metaphorical extension of a purely physical action, i.e., *tužit'sja* 'make a physical effort'. Its root is also found in *potuga*, that can mean both '(vain, unsuccessful) attempt' and 'muscular contraction', the source of force. The fact that animals (5.7%, with $z = 1.138$) and phenomena of nature (1.9%, $z = 1.155$) fulfill the subject function more often sets this verb apart from the others in the cluster. This type of attempt can be applied to physical actions that affect an "other" (22.6%, with $z = 1.175$), but also to emotions (1.9%, with $z = 1.155$). If the attempt fails, internal reasons are given (9.4%, with $z = 1.144$):

(10) Так вот и **Маруся** сейчас **тужилась улыбнуться**: *всю свою силу в одно место собрала − к губам − и не может*, вот − не может […] [Е. Замятин. На куличках]
      'So Marusja also tried (exerted herself) to smile now: she gathered all her strength to one place, towards her lips, and she can't, well, she can't […]'

4.2.3. [YOU WON'T SUCCEED]

The final cluster, amalgamated last into the overarching cluster, consists of [[*norovit'* and *poryvat'sja*] and *silit'sja*]. These verbs prefer to occur as participles (*t*'s range from 0.632 to 1.214). The infinitive actions that are attempted express a type of physical motion ($t = 0.924$) that is often not controllable ($t = 0.548$). The action can be carried out by an inanimate

subject ($t = 0.809$ for phenomena of nature and $t = 0.774$ for bodyparts) and are often repeated (t ranges from 0.678 to 1.092). If the attempt remains unsuccessful, both external ($t=0.627$) and internal ($t = 0.429$) reasons are given for the failure.

Let us turn to the within-cluster differences and take a closer look at *silit'sja*. *Silit'sja* is the verb that, more than any of the ofther verbs, prefers subclauses (44.2%, $z = 1.026$) and disprefers main clauses (55.8%, $z = -1.026$). Subjects of *silit'sja* can be abstract concepts (0.8%, $z = 1.01$). As for infinitive actions, *silit'sja* has the highest percentage of non-controllable infinitive actions (37.1% with $z = 1.153$) as well as weakly controllable infinitive actions (16.1%, $z = 1.127$) and, conversely, the lowest percentage of controllable infinitives (46.5% or $z = -1.15$); cf. (11). Like with *starat'sja*, these infinitive actions include active and passive perception (5.4% and 2.5%, z's = 1.155) and figurative extensions of putting someone else to motion (1.2%, $z = 1.151$) but can also express mental actions (38%, $z = 1.153$). Typically, the attempt fails: adverbs that stress vainness (1.7%, $z = 1.155$) or the combination of vainness and intensity (0.4%, $z = 1.155$) discriminate well between *silit'sja* and the other verbs in the cluster:

(11) Он **силился** что-то сказать, *но у него ничего не получалось*: лишь жалкое мычание доносилось *сквозь стиснутые зубы* да пузырилась в уголках рта пена … [Громов Вадим. Компромат для олигарха]
      'He tried to say something, but he couldn't: only a pitiful mumbling could be heard through his clenched teeth [...]'

The semantic differences between *silit'sja*, *starat'sja* and *probovat'* are illustrated in the following example (12):

(12) Но **Сирота** все еще **силился что-то сказать**, и *снова невозможно было понять ни слова из того, что он говорил*. Малинин наконец не выдержал и прекратил эту обоюдную муку:
      Ты не старайся, Сирота, все равно я не понимаю: *у тебя рот разбитый…* Звук и только, а голоса нет. В госпитале полежишь − восстановится, а сейчас **не пробуй**, не мучь себя […] [К. Симонов. Живые и мертвые]
      'But Sirota was still trying [*silit'sja*] to say something, and again it was impossible to understand a word of what he was saying. Finally, Malinin could not take it any longer and put an end to this mutual torture:
      You don't try [*starat'sja*], Sirota, I can't understand you anyway: your mouth is severely injured …. There is only sound, no voice.

You'll be in hospital for a while − it will heal, but for now don't try [*probovat'*], don't torture yourself [...]'

The victim has a severe mouth-injury, something he cannot possibly overcome on the spot by putting in more effort; he will have to go to hospital. However much effort the victim puts into saying something, he won't succeed, and *silit'sja* is in place. His friend, who wants to make him stop trying, uses *starat'sja* that has a comparable effort level, but can form an imperative because it lacks the association of vainness or futility *silit'sja* entails. Apart from trying to make him stop putting in effort, the friend urges the wounded man not to 'experiment' either to find a way to make himself understood.

The second node in this cluster, [*poryvat'sja* and *norovit'*], contains two verbs expressing that you are trying to do something while acting on impulse or capriciousness: *norovit'* incorporates the abstract concept *norov* 'obstinacy, capriciousness', whereas *poryvat'sja* relates to the sense 'make jerky movements' and incorporates *poryv* 'impulse'; cf. (13) and (14).

The verbs differ in more than one respect. *Poryvat'sja* expresses an attempt that can be carried out by groups or organizations (0.8%, $z = 1.005$). It favors repetition (18.5%, $z = 1.134$ for adverbs that express repetition in general) and intensification (0.8%, $z = 1.155$) and combines easily with verbs that express communication/interaction (22.7% with $z = 1.039$). These infinitive actions are often imperfective (z ranges from 1.082 to 1.154), which fits in well with the verb's preference for repetitive adverbs. If the attempt fails, external reasons are given (36.9% with $z = 1.14$):

(13)   *Несколько раз я* **порывался** *уйти и предоставить их самим себе, но меня не отпускали.* [Борис Левин. Блуждающие огни]
       'A couple of times I tried to go away and leave them to their own resorts, but they did not let me go.'

Norovit' is a type of attempt that can be initiated by phenomena of nature (6%, $z = 1.152$), bodyparts (3.2%, $z = 1.066$), man-made things (4.4%, $z = 1.155$) and animals (6.4%, $z = 1.144$). As the list of variables that are most strongly attracted to this verb point out, the verb expresses that the subject is trying to carry out an action, especially physical actions that involve the "other" (4.4% of all examples, with $z = 1.146$), actions that involve figurative physical motion (10%, with $z = 1.142$), that imply motion or the conceptually related transfer/exchange (9.2% with $z = 1.64$). Declarative sentences are avoided (96.4%, $z = -1.061$), whereas exclamative clauses are attracted (3.6%, $z = 1.061$). Attempts

are constantly repeated as indicated by *vsë vremja* ('all the time') (5.2%
with $z = 1.151$), yet do not take a very long time (1.6%, $z = -1.017$).
If the attempt fails, internal reasons are not typically given (0% with
$z = -1.147$):

(14)   Он знал, что в доме еще достаточно соли, но жена этой
        просьбой как бы привязывала его к семье, от которой, как
        ей казалось, он *все* **норовил** оторваться ради каких-то особых
        мужских или общечегемских дел. [Ф. Искандер. Сандро из
        Чегема]
        'He knew that there was enough salt in the house, but his wife
        kind of tied him with this request to the family from which he, as
        it seemed to her, continuously tried to tear himself away because
        of some specifically male or general chegemian causes.'

5. Main conclusions and prospects for further research

In this paper, we have presented the foundation for a comprehensive
and verifiable quantitative methodology designed to address the struc-
turing and description problem in research on near synonyms. More
specifically, we introduced a set of both exploratory and hypothesis-
testing statistical techniques for analyzing corpus-based behavioral pro-
files on the basis of which the internal structure of a cluster of nine near
synonymous Russian verbs was laid bare.

   Our results are largely in line with the state-of-the-art analyses of these
verbs by lexical semanticists who are native speakers of Russian. Given
the overall similarity of the traditional lexicographic analysis to our
quantitative corpus-based analysis, the *advocatus diaboli* might doubt the
utility of the proposed approach. However, this doubt would be unjusti-
fied since the present approach has several advantages over traditional
analyses.

   First, with the exception of a few semantic variables for which opera-
tional definitions were designed, the analysis is based on data that were
manually coded for a number of parameters not requiring semantic
intuitions. Thus, this behavioral profile approach is extremely precise
and explicit: (i) it takes syntactic structures more seriously than any ap-
proach with an arbitrary collocate window and (ii) it exhibits a much
lower error rate than many purely machine-based parses do. The same
holds for all other variables that were hand-coded − in fact, many of
the semantic variables we analyze defy an accurate purely automatic
coding. It is hard to see how the data could be annotated for automati-
cally, say, the semantic primitives or the literal vs. figurative distinctions

with the high degree of precision we are aiming at in this analysis. In other words, we submit that precision and recall outperform all competing approaches we are aware of.

Second, this manual annotation of largely objectively measurable characteristics has several desirable consequences. On the one hand, it avoids both the intuitive, post hoc flavor that comes with many traditional semantic analyses as well as the methodological problems associated with more introspective accounts (cf. Sandra and Rice 1995 for discussion). As a result of this, the present approach likewise allows to test the results for replicability and reliability (cf. below for some suggestions) while more intuitive analyses defy rigorous verification/falsification. By the same token, using the precise behavioral profiles, our analysis sheds light on those cases where traditional analyses and our approach yield conflicting results. For example, while our results are largely similar to the lexicographic ones, there are differences both in terms of the internal structure of the group of near synonyms and the degree to which the motivation of this structure is made explicit. That is, we do get groupings similar to previous studies, but the verbs are chained to each other in a different way all through the dendrogram, and the behavioral profiles make explicit what motivates these differences.

Finally, the present approach appears to be an interesting way to discover the formal and/or semantic clues (and, of course, their intercorrelations) that are most discriminatory within the space of formal and semantic characteristics of the set of near synonymous words. While research in first language acquisition and/or second language learning typically does not utilize exploratory data analysis methods such as cluster analysis, the present data indicate the wealth of patterns on which research into bootstrapping could be based.

The corpus-based approach we have proposed can be extended, however. The previous sections have shown that behavioral profiles can be put to use to describe similarities and differences between near synonyms on the denotational level. Yet, our tags might not provide us with all the scales of variation needed to describe the prototype center of verbs that are marked in any way since at present behavioral profiles do not include information on differences in register, style, expressiveness etc. (for a similar separate treatment of denotation, see Edmonds and Hirst 2002: 115). Obviously, it is possible to extend this analysis to include such data, yet in that case it would be of utmost importance not to neglect the variation that is induced by this factor.

Additional research is likewise needed to validate our (preliminary) results, however. Further empirical underpinning for exploratory research using clustering algorithms can be sought in different directions, of which we mention the two most frequently used. On the one hand,

one could tag larger data samples from stratified corpora for one (or some of) the parameters we have found to be relevant and subject the data to strict significance testing in order to decide whether the findings can be extrapolated to the larger population. On the other hand, and in our view more promisingly, one could use our findings to set up experiments asking native speakers to fill out questionnaires or to perform sorting tasks. One example of such an approach would be to ask subjects to sort sentences containing one of the nine near synonymous verbs into groups. This strategy appears particularly attractive since the sentences to be sorted can be formulated in a way that they foreground the parameters of interest. In addition, the results of sorting experiments lend themselves well to a cluster-analytic evaluation and, thus, to a direct comparison to the dendrogams derived from the corpus data. Methods such as those proposed by Fowlkes and Mallows (1983), for example, would be a good starting point. Another strategy would be the use of gap-filling experiments. Subjects are given concordance lines from a corpus from which the verb under consideration has been deleted and are asked to fill in the verb they think is missing. This approach allows for testing to what degree the subjects can infer the right verb from context (thus indicating the degree of discriminability of the verbs) as well as which verbs appear most similar in terms of intersubstitutability (in case the "incorrect" verb was been inserted); cf. Charles and Miller (1991) for discussion.

We hope that this paper will stimulate future research along the lines suggested, given that we have shown how rewarding behavioral profiles and the proposed methods for their evaluation are for the analysis of near synonyms in particular and lexical-semantic research in general. The type of fine-grained results we present can be interesting for scholars from a wide variety of linguistic disciplines (e. g., lexicographers, semanticists, syntacticians) as well as from different methodological backgrounds; the richness of behavioral profiles provides an ideal-starting point for research concerning interfaces between different levels of linguistic analysis, e. g., the syntax-lexis interface. In addition, results of this type may also be relevant to researchers from neighboring disciplines, such as psycholinguistics, when it comes to formulating and evaluating hypotheses concerning the interaction between grammar and lexicon in language acquisition and the mental reality of radial networks.

Notes

1. Apresjan (1995: 251−255) defines semantic fields as containing a multitude of
   meanings that have at least one semantic component in common. However, as a
   distribution test with native speakers of Russian shows (Divjak 2004: 119−127),
   the verbs *poryvat'sja* ('try, endeavor'), *tščit'sja* ('try, endeavor'), *pyžit'sja* ('go all
   out'), and *norovit'* ('try, strive to, aim at'), which Apresjan excludes from his list
   of tentative verbs, have more in common with tentative verbs than with the inten-
   tional verbs *chotet'* ('want, intend') and *namerevat'sja* ('intend, mean'), although
   it is indeed likely to assume that all of the verbs listed share at least one character-
   istic.

2. Cf. Harris (1951, 1954) as well as Rubenstein and Goodenough (1965) for early
   works; cf. Fisher, Gleitman, and Gleitman (1991), Levin (1993), and Goldberg
   (1995) on the correlation between verbs' semantics and their subcategorization
   patterns and Miller and Charles (1991), Lin (1998), and Schulte im Walde (2000)
   for psycholinguistic and corpus-linguistic examples; finally, cf. Mc Donald (1997)
   for an evaluation of corpus-derived measures of semantic similarity.

3. The translations of the verbs are taken from the Russian-English Oxford Dictio-
   nary (1993); many native speakers of English who have an advanced command
   of Russian have pointed out that these translations neither fully capture the rich-
   ness of the Russian counterparts nor sound natural in English. Following their
   suggestion, we translate all examples using *try*, providing the more specific dictio-
   nary solution between brackets. Benson and Benson (1995) mention only 'try' for
   probovat' and add 'try' to *norovit'*, which we will take over here. Given that Eng-
   lish does not have lexemes that adequately cover the contents of the Russian
   verbs, translations will not be provided in the remainder of this article.

4. The elicitation experiment with native speakers of Russian was set up as a small
   number design. Five native speakers between the ages of 25 and 50 judged the
   constructional possibilities of 300 verbs on a three-point scale. The experiment
   was conducted over a period of three months in the form of a weekly interview
   during which each native speaker was presented with approximately 25 verbs in
   the 3 different constructions tested for. Several measures were taken to minimize
   the obvious negative effects of this set-up. Native speakers were asked both to
   judge ready made sentences and to form sentences using particular constructional
   devices; these sentences were on a later occasion presented to the participant who
   had constructed them as well as to other participants. To guard against lexical
   effects, the tests were carried using pro-nouns and other pro-forms (cf. Smessaert
   et al. 2005), which ensures that the mutual effect of lexical items in a construction
   is minimized as much as possible; as a result, the acceptability or unacceptability
   of a particular construction is very unlikely to be influenced by a particular lexical
   compatibility or incompatibility of words that are not focused on. In the experi-
   ment, infinitives were replaced by *(s)delat'* ('do'), and the results might therefore
   be restricted to combinations with infinitives that relate to this action primitive.

Moreover, to check for repetition effects in judgments of grammaticality 10 control judgments were collected for every verb in every construction type from an ever varying pool of native speakers. In this case, the trigger questions were mixed with other, non-related questions about aspects of Russian syntax and semantics. Finally, the results obtained have been systematically compared against information contained in dictionaries, and utterances found in the Amsterdam Corpus and on the Web. Properties that are stated to be absent were rejected by native speakers and not found in the corpus or on the Web. For a more detailed discussion of the data collection and native speaker survey we refer to Divjak (2004: 19−33).

5. Verbs, the constructional possibilities of which are limited to combinations with an infinitive, lack conspicuous formal characteristics that can be used as the basis for further subcategorization. Therefore, these verbs were classified with the help of elicited and non-elicited data on their linear distribution, i. e., their mutual combinatorial possibilities in verb triples. This procedure yielded categories that have received formal underpinning from implication relations and aspectual behavior (Divjak 2004).

6. Reduction to a relatively representative sample was carried out in two steps. First, all examples for all nine verbs available in the corpus were called up. Next, the total number of examples was reduced to approximately 250 by selecting every n-th example. Although 250 may seem a small number for lexical semantic analysis, it is a number for which detailed manual tagging is practically feasible and it is more than double the amount of the 100 examples per verb used in the most extensive study of (four of) these verbs available (Apresjan et al. 1999: 303−308). The current samples were manually post-edited to reach an acceptable degree of representativity and contain example sentences drawn from different genres and works from (on average) 75 authors, no author providing more than five percent of all examples for a verb, no single text providing more than 3 examples. For some verbs, *tščit'sja*, *pyžit'sja* and *tužit'sja*, the combined corpora did not contain enough examples, and we supplemented corpus data with internet data. We are aware of the fact that this is not the optimal procedure since, among other things, data drawn from the internet are possibly not well-balanced as search engines do not provide a way of arriving at an equal representation of different genres, styles, and possible theme-related preferences. However, as Kilgarriff and Grefenstette (2003: Section 4) convincingly argue, it is difficult, if not impossible, to decide for any general purpose corpus whether it is representative for a particular language. Moreover, Keller and Lapata (2003) show that frequency counts for rare bigrams from the web are comparable with counts based on the North American News Text Corpus (350m words). For these reasons we believe that our strategy of supplementing corpus data with additional data from the web is licensed. Moreover, we tried to reduce the kind of register-specific effects mentioned above further by restricting internet searches with the search engine Yandex to literary pages for *tščit'sja*. For *pyžit'sja* and *tužit'sja,* verbs that belong to colloquial Russian, this was not an option and non-literary examples were taken into account. Yet, all examples were checked manually to ensure that they are not related by author or theme. In addition, the acceptability of the examples has been tested by presenting them to native speakers. Constructions uninterpretable to the participants were weeded out. Thus, while we admit that our results should be checked against data from stratified corpora once they become available, we submit that our retrieval strategy is as good as is possible, given the sparseness of the Russian verbs investigated here.

7. To ensure reliability of the tags, the whole sample was tagged twice and the elements that were assigned the same tag have been compared groupwise.

8. This conclusion is in line with contemporary Russian views on the status of, e. g., adverbs. Chrakovskij (1998: 152) pleads for excluding adverbs from grammar: specific verbs have clear preferences for specific adverbs. Therefore, information about "circumstants" should be considered lexical and thus listed in dictionaries, like information about "actants". The solution is an "integral description of language" proposed by Apresjan et al. (1995) where lexicon and grammar interact. Yet, though Apresjan et al. (1999: 303−308) include a description of grammatical phenomena in their dictionary, the grammatical part of the descriptions offered is not related to the detailed lexical information provided.

9. In order to ascertain that the absence of these three basic areas of human behavior in our dataset was not due to sampling or coding errors, a small-scale experiment with five native speakers of Russian was conducted; we found that physiological reactions and conditions as well as wishes/desires were dispreferred with tentative verbs.

10. Edmonds and Hirst (2002: 125) argue that there are different methods to compute the degree of semantic similarity between any two words. Examples they give include using taxonomic lexical hierarchies or lexical semantic networks, large text corpora, word-definitions in machine-readable dictionaries etc. However, these measures are respectively inappropriate, not fine-grained enough or not capable of uncovering the more subtle differences in the use of near synonyms. Edmonds and Hirst do however suggest (2002: 118 n. 13) using cluster analysis for the automatic discovery of near synonym clusters; for a general overview of such clustering applications; cf. Manning and Schütze (2000: Chapter 14).

11. As the term "(dis)similarity" indicates, in the context of hierarchical agglomerative cluster analysis, measures of similarity and measures of distance are often understood as complementary and directly derivable from each other, and we will use *similarity* and *distance* in exactly this way.

12. In addition, other measures such as Euclidean distance and Weighted pair group average are wanting given that our dataset does not fulfill the requirements these algorithms impose on the data to be processed. Among other things, the minimal Euclidean distance is sensitive to outliers and it is only advised to use the Weighted pair group average when the clusters are suspected to be greatly uneven.

13. While some of the absolute numbers with which some of these variables occur are small, they still have high predictive power, given that the *t*-values and the *z*-sores take into consideration the variable's dispersion across all clusters (in the case of *t*-values) and verbs (in the case of *z*-scores).
For the sake of completeness, we have also computed *F*-values to assess the clusters' homogeneity. The numbers of *F*-values larger than 1 are 40, 24, and 29 for [*probovat'*, *pytat'sja*, *starat'sja*], [*silit'sja*, *poryvat'sja*, *norovit'*], and [*tščit'sja*, *pyžit'sja*, *tužit'sja*] respectively. On the basis of *z*-scores it can be determined where the inhomogeneity of the clusters arises most strongly, cf. below.

14. Note that while the two clustering algorithms agree as to the overall structure in the dendrogram, they differ with respect to the internal structure of one cluster: The HAC outputs [[*tščit'sja* and *pyžit'sja*] and *tužitsja*] whereas Fitch outputs [[*pyžit'sja* and *tužitsja*] and *tščit'sja*]. At present, we are not able to relate this difference to any semantic properties of the three verbs and attribute it to the mathematical properties of the clustering algorithms. In the presentation, we will follow the solution proposed by the hierarchical cluster analysis, on the basis of which the *t*-values and *z*-scores are computed.

15. Concepts can be categorized on different levels of a taxonomy. It is hard to determine how many levels exactly need to be distinguished in a hierarchical system, the more because the number of levels may be influenced by personal experience:

familiarity may lower the basic level which in turn affects the structure and contents of the sub- and super-ordinate levels (Rosch et al. 1976: 430). Edmonds and Hirst (2002: 115) have found that it is easier to model near synonyms if the fine-grained differences are accounted for on the subconceptual level: "a 'clustered model of lexical knowledge' is most suited for representing near synonyms. Near synonyms are so closely related in meaning that they share all coarse-grained aspects and differ only in fine-grained aspects that are most effectively modeled at the subconceptual level". This does not imply, however, that these differences are located on the subconceptual level in (psychological) reality.

16. A Fisher-Yates exact test for contingency shows that the preferences of *pytat'sja* and *starat'sja* for repetition and intensity respectively are significant ($p = 0.037$, odds ratio = 5.62).

17. Krasnych (1979: 79) reports a similar preference of *pytat'sja* and *starat'sja* for occurring in the imperfective ($\chi^2 = 8.05$; *df* = 1; $p < 0.001$ − statistics computed from Krasnych's reported frequencies).

References

Apresjan, Jurij Derenikovič
1995a   *Избранные Труды. Том I. Лексическая Семантика: Синонимические Средства Языка.* Moskva: Škola "Jazyki Russkoj Kul'tury".
1995b   *Избранные Труды. Том II. Интегральное Описание Языка и Системная Лексикография.* Moskva: Škola "Jazyki Russkoj Kul'tury".

Apresjan, Jurij Derenikovič et al.
1995   *Новый Объяснительный Словарь Синонимов: Проспект.* Moskva: Russkie Slovari.
1999   *Новый Объяснительный Словарь Синонимов Русского Языка. Vol. I.* 2nd ed. Moskva: Škola "Jazyki Russkoj Kul'tury".
2000   *Новый Объяснительный Словарь Синонимов Русского Языка. Vol. II.* Moskva: Škola "Jazyki Russkoj Kul'tury".

Arppe, Antti and Juhani Järvikivi
2002   Verbal synonymy in practice: combining corpus-based and psycholinguistics evidence. Poster at the Workshop Quantitative Investigations in Linguistics at the QITL2 in Osnabrück, Germany.

Atkins, Beryl T. S.
1987   Semantic ID tags: corpus evidence for dictionary senses. *Proceedings of the Third Annual Conference of the UW Centre for the New Oxford English Dictionary*, 17−36.

Atkins, Beryl T. S. and Beth Levin
1995   Building on a corpus: a linguistic and lexicographical look at some near synonyms. *International Journal of Lexicography* 8, 285−114.

Backhaus, Klaus et al.
1996   *Multivariate Analysemethoden: eine anwendungsorientierte Einführung.* 8. ed. Berlin, Heidelberg, New York: Springer.

Barentsen, Adriaan A. (Compiler).
1999   The Amsterdam Corpus of Written Russian.

Benson, Morton and Eveline Benson
1995   *Русско-Английский Словарь Глагольных Словосочетаний.* Moskva: Meždunarodnaja Škola Perevodčikov.

Ways of trying in Russian: clustering behavioral profiles   57

Bricyn, V. M.
    1990   *Синтаксис и Семантика Инфинитива в Современном Русском Языке.* Kiev: Naukova Dumka.

Chrakovskij, Viktor S.
    1998   Понятие сирконстанта и его статус. *Семиотика и Информатика* 36: 141−153.

Cruse, D. Alan
    1986   *Lexical Semantics.* Cambridge: University Press.

Černova, S. V.
    1996   Модальные глаголы в современном русском языке. *Семантическая Модель 'Замысел − Осуществление Замысла'.* Kirov: Izdatel'stvo VGPU.

De Jonge, Bob
    1993   The existence of synonymy in a language: two forms but one, or rather, two meanings? *Linguistics* 31(3), 521−538.

Divjak, Dagmar
    2003   On trying in Russian: A tentative network model for near(er) synonyms. In: *Slavica Gandensia* 30. *Belgian Contributions to the 13th International Congress of Slavicists, Ljubljana, 15−21 August 2003*, 25−58.
    2004   Degrees of verb integration. Conceptualizing and categorizing events in Russian. Unpublished Ph.D. Dissertation, K.U. Leuven (Belgium), Dept. of Oriental and Slavic Studies.
    2006   Ways of intending: A corpus-based cognitive linguistic approach to near synonyms in Russian. In: Gries, Stefan Th. and Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis.* Berlin and New York: Mouton de Gruyter, 19−56.

Edmonds, Philip and Graeme Hirst
    2002   Near synonymy and lexical choice. *Computational Linguistics* 28(2), 105−144.

Evgen'eva, A. P. (ed.)
    2001   *Словарь синонимов русского языка в двух томах.* 2nd ed. Moskva: Astrel' AST.

Felsenstein, Joe
    2005   Phylogeny Inference Package PHYLIP 3.65, distributed by the author, http://www.phylip.com.

Fisher, Cynthia, Henry Gleitman, and Lila R. Gleitman
    1991   On the semantic content of subcategorization frames. *Cognitive Psychology* 2 (3), 331−392.

Fowlkes, Edward B. and Colin L. Mallows
    1983   A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association* 78 (383), 553−569.

Geeraerts, Dirk
    1985   Preponderantieverschillen bij bijna-synoniemen. *De Nieuwe Taalgids* 78, 18−27.
    1988   Where does prototypicality come from? In: Rudzka-Ostyn, Brygida (ed.), *Topics in Cognitive Linguistics.* Amsterdam/Philadelphia: John Benjamins, 207−229.

Goldberg, Adele E.
    1995   *Constructions: a Construction Grammar Approach to Argument Structure.* Chicago and London: Chicago University Press.

Gries, Stefan Th.
    2001   A corpus-linguistic analysis of -ic and -ical adjectives. *ICAME Journal* 25: 65−108.

2003   Testing the sub-test: A collocational overlap analysis of -*ic* and -*ical* adjectives. *International Journal of Corpus Linguistics* 8(1), 31−61.

2006   Corpus-based methods and cognitive semantics: The many senses of *to run*. In: Gries, Stefan Th. and Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin and New York: Mouton de Gruyter, 57−98.

Gries, Stefan Th., Beate Hampe, and Doris Schönefeld
2005   Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4), 635−676.

Haiman, John
1983   Iconic and Economic Motivation. *Language* 59(4), 781−819.

Hanks, Patrick
1996   Contextual dependency and lexical sets. *International Journal of Corpus Linguistics* 1 (1), 75−98.

Harris, Zellig S.
1951   *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
1954   Distributional structure. *Word* 10 (2−3), 146−162.

Hirst, Graeme
1995   Near synonymy and the structure of lexical knowledge. In: *Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity. Papers from the AAAI Spring Symposium*, 51−56. Menlo Park: AAAI.

Janda, Laura A.
2004   A metaphor in search of a source domain: The categories of slavic aspect. *Cognitive Linguistics* 15(4), 471−527.

Kaufman, Leonard and Peter J. Rousseeuw
1990   *Finding Groups in Data*. New York: John Wiley.

Keller, Frank and Mirella Lapata
2003   Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics* 29 (3), 459−484.

Kilgarriff, Adam and Gregory Grefenstette
2003   Introduction to the special issue on the web as corpus. *Computational Linguistics* 29 (3), 333−347.

Krasnych, V. I.
1979   О наиболее употребительных глаголах, сочетающихся с инфинитивом. In: *Русский Язык для Студентов-Иностранцев. Сборник Методологических Статей*. Moskva: Russkij Jazyk, 74−86.

Langacker, Ronald W.
1987   *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press.

Levin, Beth
1993   *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Li, Hang and Naoki Abe
1998   Word clustering and disambiguation based on co-occurrence data. *Proceedings of COLING 17*, 749−755.

Lin, Dekang
1998   An information-theoretic definition of similarity. *Proceedings from the 15th International Conference on Machine Learning*, 296−304.

Manning, Christopher D. and Hinrich Schütze
2000   *Foundations of Statistical Natural Language Processing*. Fourth printing with corrections. Cambridge, MA: The M.I.T. Press.

Mangasser-Wahl, Martina (ed.)
  2000   *Prototypentheorie in der Linguistik: Anwendungsbeispiele − Methoden −
            Reflexion − Perspektiven.* Tübingen: Stauffenburg.
McDonald, Scott
  1997   Exploring the validity of corpus-derived measures of semantic similarity.
            Paper presented at the 9th Annual CCS/HCRC.
Miller, George. A. and Walter G. Charles
  1991   Contextual correlates of semantic similarity. *Language and Cognitive
            Processes* 6(1), 1−28.
Mondry, Henrietta and John R. Taylor
  1992   On lying in Russian. *Language and Communication* 12(2), 133−43.
Ožegov, Sergej I. and Natal'ja Ju. Švedova
  1999   *Толковый словарь русского языка.* Moskva: Azbukovnik.
The Oxford Russian Dictionary
  1993   Edited by P. Falla, M. Wheeler, and B. Unbegaun. Oxford: Oxford Uni-
            versity Press.
Pearce, Darren
  1992   Synonymy in collocation extraction. *Proceedings of the NAACL 2001
            Workshop on WordNet and Other Lexical Resources*, 41−46.
R Development Core Team
  2005   *R 2.2 − A language and environment for statistical computing. Vienna: R
            Foundation for Statistical Computing.* ISBN 3-900051-07-0, http://
            www.R-project.org.
Rachilina, Ekaterina V.
  2000   *Когнитивный Анализ Предметных имен: Семантика и Сочетае-
            мость.* Moskva: Russkie Slovari.
Rosch, Eleanor, Caroline B. Mervis, Wayne Gray, David Johnson, and Penny
    Boyes-Braem
  1976   Basic objects in natural categories. *Cognitive Psychology* 8(4), 382−439.
Rubenstein, Herbert and John B. Goodenough
  1965   Contextual correlates of synonymy. *Communications of the ACM* 8(10),
            627−633.
Sandra, Dominiek and Sally Rice
  1995   Network analyses of prepositional meaning: mirroring whose mind −
            the linguist's or the language user's? *Cognitive Linguistics* 6(1), 89−130.
Schulte im Walde, Sabine
  2000   Clustering verbs semantically according to their alternation behaviour.
            *Proceedings of the 18th International Conference on Computational Lin-
            guistics*, 747−753.
Smessaert, Hans, Bert Cornillie, Dagmar Divjak and Karel van den Eynde
  2005   Degrees of clause integration. From endotactic to exotactic subordina-
            tion in Dutch. *Linguistics* 43(3), 471−530.
Stefanowitsch, Anatol and Stefan Th. Gries
  2003   Collostructions: investigating the interaction between words and con-
            structions. *International Journal of Corpus Linguistics* 8 (2), 209−243.
Taylor, John R.
  2003   Near synonyms as co-extensive categories: high and tall revisited. *Lan-
            guage Sciences* 25 (3), 263−284.
Turney, Peter D.
  2001   Mining the web for synonyms: PMI-IR versus LSA on TOEFL. *Proceed-
            ings of the 12th European Conference on Machine Learning*, 491−502.

Vanhatalo, Ulla
    2003   Evaluating the semantic content of near synonyms: population tests versus corpus linguistics. *Virittäjä* 3, 107.
Vendler, Zeno
    1967   Verbs and times. In: Vendler, Zeno (ed.), *Linguistics in Philosophy.* New York: Ithaca, 97−121.
Ward, Joe H. Jr.
    1963   Hierarchical groupings to optimize an objective function. *Journal of the American Statistical Association* 58 (301), 236−244.
Wierzbicka, Anna
    1988   *The Semantics of Grammar.* Amsterdam/Philadelphia: John Benjamins.
    1996   *Semantics: Primes and Universals.* Oxford: Oxford University Press.
    1990   Toward a theory of syntactic prototypes. In: Tsohatzidis, Savas L. (ed.). *Meanings and Prototypes: Studies in Linguistic Categorization.* London, New York: Routledge, 285−306.