

Dmitrij O. Dobrovol'skij

Are free word combinations really free?

Lexical co-occurrence of the Russian degree modifier *črezvyčajno*¹

Abstract: The linguistic focus of the present paper is on the nature of usage constraints. The leading hypothesis is that there exists a kind of lexical co-occurrence which takes an intermediate position between the completely arbitrary and the strictly predictable, i.e. between compositionality and phraseology. This “intermediate” type of co-occurrence has to be described in terms of tendencies or principles which mainly operate within certain lexical classes. Describing lexical items of this type is a challenge, as it is necessary to develop metalinguistic instruments that would allow the determination of selection restrictions which can neither be traced back to the semantics of a given lexical item nor be fixed as a strictly limited list of restricted collocations.

The Russian degree modifier *чрезвычайно* (*črezvyčajno*), which literally means ‘extraordinarily’, was taken as the starting point for the empirical analysis. The aim of the analysis is to discover and exemplify phenomena of lexical co-occurrence which are neither phraseologically bound nor absolutely free. This word seems to be especially suitable for studying the phenomena in question because it reveals some selection restrictions which are hard to predict.

In addition, the present paper focuses on how corpus evidence relates to survey data. First, using corpus-based techniques, data were obtained concerning the lexical semantics and usage conditions of the Russian degree modifier *črezvyčajno*. These data were then compared to the results of a survey. The results show a strong convergence between these two methods of analysis. Usage conditions and the relevant restrictions of the word *črezvyčajno* are based, to a great extent, on the native speaker's competence. Whether or not corpus data reflects this kind of competence is an issue which is not self-explanatory and deserves special attention.

¹ This paper is based on work supported by the RGNF under Grant 11-04-00105a, and by the Basic Research Program “Corpus Linguistics” of the Presidium of the Russian Academy of Sciences. I am most grateful to Ju.D. Apresjan, E.V. Padučeva and E.V. Uryson for sharing their ideas with me about some issues discussed in this paper, to F. Čermák, K. Dill, D. Divjak, A. Dziemianko, St.Th. Gries, St. Miller and my anonymous reviewers for their suggestions for improvement, and to D.V. Ljusin who helped me with the statistical analysis of the survey data.

Keywords: selection restrictions, collocation, degree modifier, combinatorial profile, Meaning \leftrightarrow Text Theory, Lexical Functions, Construction Grammar, corpus evidence, survey data

Correspondence address: dobrovolskij@gmail.com

1 Aims and assumptions

The general aim of this investigation is to reveal the nature of the grey zone between the two extremes of lexical co-occurrence: strict predictability and complete arbitrariness.² The study is based on the idea that many phenomena of lexical co-occurrence traditionally regarded as arbitrary are actually motivated (cf. similar ideas in Apresjan 2004), not in the sense that they are fully predictable and deducible from the semantic definitions of the combining lexemes, but in the sense that they are guided by certain principles.

The Russian degree adverb *črezvyčajno* (literally ‘extraordinarily’) provides the empirical basis of this study. The specifics of this word’s combinatorial profile make it especially suitable for testing assumptions concerning tendencies or principles governing lexical co-occurrence. Corpora and surveys are excellent sources for analyzing the real usage of this word.

Within the theoretical framework of the Meaning \leftrightarrow Text Theory, the degree modifier *črezvyčajno* represents a value of the Lexical Function Magn [“magnifier”] (in the sense of the Meaning \leftrightarrow Text Theory), i.e. it is an intensifier.

The metalanguage of Lexical Functions is perhaps the most well-known part of the Meaning \leftrightarrow Text Theory. Research has shown that it is a powerful instrument for describing the co-occurrence behaviour of lexemes, and especially relevant selection restrictions which seem to be unpredictable. Lexical Functions were first introduced in Žolkovskij and Mel’čuk (1965, 1967), primarily to formalize the description of collocations. Lexical Functions were also used for the representation of semantic derivatives (see for detail Mel’čuk 2007). But this aspect can be left aside here. The argument of a Lexical Function is the keyword (the base of the collocation); the value of the function is the collocate of the collocation containing this keyword. One of the conditions of applying Lexical Functions to

² The term *co-occurrence* is used here in the sense of a capacity of a given lexical item to occur alongside with certain contextual partners, rather than in the sense of the above-chance occurrence of two elements.

collocation analysis is that the value of a given function (the collocate) has to be more or less bound by its keyword.

For example, the verbal Lexical Functions Oper_1 and Oper_2 can be illustrated by the following collocations with the base *support*: *to lend support* [$\text{Oper}_1(\text{support } N) = [\text{to}] \text{ lend } [\sim \text{ to } N]$] and *to receive support* [$\text{Oper}_2(\text{support } N) = [\text{to}] \text{ receive } [\sim \text{ from } N]$]. The Lexical Function Magn, which plays a central role in this paper, means ‘very’, ‘to a (very) high degree’, ‘intense(ly)’. So, Magn from Russian *dožd* ‘rain’ is the adjective *prolivnoj* ‘pouring’, Magn from Russian *brjunet* ‘person with black hair, brunette’ is the adjective *žgučij* ‘burning’, i.e. the sense ‘person with very black hair’ is expressed in Russian by the collocation *žgučij brjunet* (lit. ‘burning brunette’). Compare as well some English examples from (Mel’čuk 2007):

Magn(*naked*) = *stark*

Magn(*patience*) = *infinite*

Magn(*skinny* [person]) = *as a rake*.

As we can see from the last example, this Lexical Function is not necessarily expressed by adjectives or adverbs.

The practical application of this linguistic tool sometimes creates certain problems: It is often not clear to what extent the collocate must be bound by the base. We are dealing with a clear case only if a given keyword demands a unique collocate, cf. the Russian collocation *prolivnoj dožd* ‘pouring rain’, where *prolivnoj* expresses the idea of ‘very intense’ only in connection with *dožd*. However, very often we encounter cases where the same keyword can be intensified by many other synonymous collocates, and/or the same collocate can be used to modify many different keywords. Compare the following example used by Mel’čuk in (2007: 130) to illustrate how the metalanguage of Lexical Functions can systematically describe the combinatorial profile of a given word. Just to name a few, the English noun *revulsion* combines with the adjectives *deep*, *extreme*, *utmost* (Magn) and *slight* (AntiMagn) and with the verbs *experience* and *feel* (Oper_1). This example shows that the Lexical Function Oper_1 for this noun can be expressed by more than one modifier, and also that each of these verbs combines with more words than just *revulsion*. The same goes for Magn. Of course, the adjectives *deep*, *extreme* and *utmost* possess very broad combinatorial profiles. But, on the other hand, not every noun can be intensified by these adjectives. In this sense, they also can be considered phraseologically bound to a certain extent.

Since one of the aims of the Lexical Functions component of the Meaning \leftrightarrow Text Theory is to obtain an exact and formalized description of combinatorial properties of every word whose co-occurrences are not fully predictable, the question arises as to how all relevant co-occurrences can possibly be registered. If we assume that every word combination of this type is quite arbitrary, such that the set of possible word combinations cannot be predicted in principle, the only way

to describe these phenomena would be to produce exhaustive lists of collocations. In cases such as [*deep* N] or [*extreme* N], this is practically impossible. It also seems neither possible nor reasonable to try to produce an exhaustive list of collocations with *črezvyčajno*. However, if we start from the assumption that there are certain tendencies or principles governing the choice of possible combinatorial partners,³ the task is to find out these principles. Principles or tendencies of this kind surely cannot guarantee the absolute predictability of every contextual partner of a given word, but it can help to formulate reasonable lexicographic expectations (in the sense of Apresjan, cf., for instance, Apresjan 2008). As for an appropriate lexicographic format of the use of *črezvyčajno*, it is not yet quite clear what it will look like in detail. It is likely that besides the meaning explication, the representation of this adverb in an active dictionary has to include comments on its definition as well as an indication of relevant semantic and syntactic constraints.

Similar goals (at least, with regard to lexicographic applications) are being pursued by FrameNet and Construction Grammar; cf. (Fillmore 2008). Construction Grammar focuses above all on linguistic phenomena which are “not strictly predictable from the properties of their component parts or from other previously established constructions” (Goldberg 1995: 4). From this perspective, all possible lexical co-occurrences which are not strictly predictable from the form or meaning of their components are the subject of Construction Grammar. “A construction is a schema or template, which captures what is common to a range of expressions, and which, at the same time, sanctions the creation of new expressions of the respective type” (Taylor 1998: 177).

In the case of degree modifiers such as *črezvyčajno*, we are dealing with exactly this kind of phenomenon; its co-occurrences are not strictly predictable. However, they are also not phraseologically bound in the sense that there is a clearly restricted list of possible contextual partners. The combinatorial potential of such words lies between “bound” and “free”. The Internet search machine Yandex (Internet address: <http://yandex.ru/yandsearch>) offers about 27 million instances of *črezvyčajno* occurring with a rather wide variety of lexical partners, so it is difficult to call it phraseologically bound. The list of possible word combinations is open, so that “new expressions of the respective type” (as Taylor expressed it) can be created, but these expressions must follow certain principles.

The metalinguistic apparatus of Lexical Functions turns out to be very useful for a systematic description of co-occurrences which can then be represented as a check list. However, if it is hardly possible to list them exhaustively, then it is not

³ In this paper, the terms *combinatorial partners*, *contextual partners* and *lexical partners* are used as synonyms to describe elements of a co-occurrence.

reasonable to use this apparatus. On the purely theoretical level, using the ideas of Construction Grammar might be more promising. As the impetus of Construction Grammar is to develop methods of analysis that would apply to both free (i.e. regular, compositional) and bound (i.e. irregular, unpredictable or idiomatic) constructions, such methods and instruments should also cover all possible “intermediate” types of co-occurrence. However, it is not quite clear (at least, at the present development stage of Construction Grammar) how such instruments should be designed in every single case. If we take the construction [črezyčajno P] as the starting point, the task of analysis is to clarify which principles govern the choice of P (P stands for the lexical item in the scope of the discussed degree modifier). But this task does not differ from simply looking for and describing all relevant combinatorial properties of *črezyčajno* itself, i.e. from formulating reasonable lexicographic expectations for this word.

Here we are dealing with “selectively compositional” phrases (cf. for this notion Kuiper [2009: 20–21]), i.e. with semi-bound phenomena. According to Mel’čuk (in this volume), compositionality displays three dimensions in compliance with three dimensions of a linguistic sign: signified, signifier, and syntactics. The construction [črezyčajno P] is compositional according to the first two dimensions, but not according to syntactics.

The standard, compositional co-occurrence of lexemes can normally be described in terms of the relevant semantic features of these lexemes and the interaction rules of the given semantic features. The problem here is that the semantic structures of near-synonymous adverbs (cf. *črezyčajno* ‘extraordinarily’, *neobyknovenno* ≈ ‘exceptionally’, *neobyčajno* ≈ ‘immensely’, *krajne* ‘utterly’, *beskonečno* ‘infinitely’, *bezмерно* ≈ ‘beyond all measure’) do not at first glance reveal any relevant distinctive features which could be regarded as factors motivating selection constraints. In fact, how can we explain that certain words prefer to combine with *črezyčajno*, others take instead *krajne* ‘utterly’ as an intensifier, some take *neimoverno* ‘incredibly’, and others *beskonečno* ‘infinitely’? Why can we use both *krajne malo* literally ‘utterly little’ and *črezyčajno malo* literally ‘extraordinarily little’ without restriction, but not **krajne mnogo* ‘utterly much’,⁴ whereas *črezyčajno mnogo* ‘extraordinarily much’ is again quite acceptable? The

⁴ In all texts contained in the Russian National Corpus, only 5 examples of *krajne mnogo* could be found, all of which sound rather odd for well-educated native speakers. However, this restriction concerning the co-occurrence *krajne mnogo* is valid maybe only for stylistically elevated contexts. In Internet blogs this co-occurrence is rather frequent. There are about 226 000 Yandex-hits, which seems to be a lot. On the other hand, co-occurrences in stylistically elevated sources are even more frequent: for *črezyčajno mnogo* there are about 698 000 hits, and for *krajne malo* about 3 million hits.

simplest way to answer these questions would be an explanation based on irregular and unpredictable usage conventions. Another possibility is to look for certain motivating factors, which do not have to possess predicting power or to go back to the differences in the semantic definitions of the words in question, but provide a kind of *post factum* explanation. If such factors could be found it would mean that the combinatorial behaviour of degree modifiers mentioned above is not arbitrary and can be described in terms of semantic (and maybe also stylistic or discursive) correlations. In other words, it would mean that there are some tendencies or principles typical of the choice of P in the construction [degree adverb_{Magn} P], or, in our case, [*črezvyčajno* P]. It is important that the semantic structure of the degree adverb in question does not need to be responsible for the selection constraints. The reason may lie in some specific properties of its lexical partner, i.e. of P.

My basic assumption is that the combinatorial profile of every word has a core and a periphery. Some lexical co-occurrences are perceived as fully corresponding to the usage norm and stylistically irreproachable, others as being erroneous, and some as being in the middle, that is to say more or less acceptable. The first ones are considered to be central, the second ones are ruled out by the usage norm; those in-between form a broad and vague periphery of the combinatorial profile. The latter part of the combinatorial profile is especially important from the point of view of linguistic change. It is not typical for a given word to abruptly change its co-occurrence behaviour. What seems to be more probable is the hypothesis that the peripheral co-occurrences of a given word can change status over time. Some can change from periphery to centre, whereas others can be excluded from the domain of normative acceptance. In certain cases, combinatorial changes of this kind can lead to the modification of the semantic structure of a given lexical item.

Corpus analysis (see Section 3 for more detail) clearly shows that, in the 19th century, the degree modifier *črezvyčajno* had a broader combinatorial profile than it has today. Cf. co-occurrences such as *črezvyčajno vykatyvala glaza* literally '[she] extraordinarily rolled out her eyes', *črezvyčajno raskryv svoj rot* 'extraordinarily opening his mouth', *lestnica byla črezvyčajno grjaznaja* 'the stairs were extraordinarily dirty' encountered in Dostoevskij's prose.

In this study, the Russian National Corpus (<http://www.ruscorpora.ru>) is used. At present, it consists of ca. 193 million running words. The corpus includes texts of various genres: fiction (about 86 million running words), social and political journalism (69.7 million), as well as educational scientific texts, letters, official business texts, etc. For more detail see <http://www.ruscorpora.ru/corpora-stat.html>. The total number of occurrences of *črezvyčajno* in the Russian National Corpus (RNC) is 13 956.

The analysis of present-day texts from RNC (both fiction and journalism) reveals the following tendencies which are typical for the contemporary usage of the word *črezyčajno*.

1. The degree modifier *črezyčajno* combines with adjectives and adverbs rather than with verbs. Co-occurrences such as *črezyčajno interesnyj/zainteresovan* ‘extraordinarily interesting/interested’ sound better to native speakers than, for example, *črezyčajno interesujus* ‘I extraordinarily interest myself’.
2. Words in the scope of *črezyčajno* more often denote non-physical, abstract properties than physical ones (*črezyčajno tjažěloe položenie* ‘an extraordinarily difficult situation’—literally ‘extraordinarily heavy situation’ is better than *črezyčajno tjažělyj čemodan* ‘an extraordinarily heavy suitcase’).
3. Words in the scope of *črezyčajno* tend to denote features which are biased towards the positive pole of various scales. For example, co-occurrences like *črezyčajno širokij/odarěnnij/ěnergičnyj* ‘extraordinarily broad/gifted/energetic’ are encountered more often than co-occurrences such as *črezyčajno uzkij/bezdarnyj/vjalyj* ‘extraordinarily narrow/ungifted/ inert’.

Similar ideas were developed in Dobrovol’skij (2005a, 2005b). In the formulation of the hypothesis introspection was also addressed, which is a fully acceptable procedure even within empirically oriented research traditions; compare, e.g., (Gries 2002). Obviously, the principles or tendencies formulated above should not be interpreted as strict rules, for technical reasons they are referred to as Principle 1, Principle 2 and Principle 3 below. These principles can be quantified on the basis of corpus analysis, i.e. they are able to be empirically supported.

Before discussing each of these principles in some detail, I would like to point out one more specific property of present-day contexts in which the usage of *črezyčajno* contrasts with contexts from the 19th century. Currently, the adverb *črezyčajno* is perceived as a word which is bound to certain thematic domains and to a discourse type which can be labelled “formal discourse” (in contrast to everyday language). For the purposes of this paper, such words will be called genre-restricted words.

2 Combinatorial profile of *črezyčajno*

2.1 Categorical and subcategorical selection constraints

Let us consider the first tendency mentioned above, Principle 1. In a sample of 500 occurrences taken from RNC-texts of the 20th–21st century, only 50 hits contain the

co-occurrence of *črezvyčajno* with verbs; in 367 hits this degree modifier combines with adjectives and participles, and in 83 cases with adverbs. The fact that degree modifiers combine with adjectives and adverbs rather than verbs is well-known. This combinatorial property was pointed out by A.M. Peškovskij (2001 [1914]: 101). According to corpus data, the word *črezvyčajno* differs from other degree modifiers (e.g., from *očen'*) by its much stronger avoidance of co-occurrences with verbs.

My analysis shows that the co-occurrence of *črezvyčajno* with verbs of certain semantic classes is in principle impossible, for example, with verbs denoting position in space, parameter or existence, such as *stand*, *sit*, *contain*, *number*, *be*. This is obvious when considering the semantics of these verbs, because such verbs cannot occur in the scope of a degree modifier. However, these natural constraints do not explain the relevant selection peculiarities of *črezvyčajno* seen in text corpora. The combinatorial profile of this word in the domain of verbs differs from that of other degree modifiers, such as *očen'*; compare *ja očen' spešil/prosil* literally 'I was very in a hurry, I very asked', but *ja črezvyčajno spešil/prosil* 'I was extraordinarily in a hurry, I extraordinarily asked' (In Russian we are dealing with the same VP-construction in both cases, namely [degree adverb_{Magn} V_{Past}]). Compare the relevant RNC data: *kto-l. očen' spešil* 'sb. was very in a hurry' (61 hits), *kto-l. očen' prosil* 'sb. very asked' (90 hits), whereas the utterances *kto-l. črezvyčajno spešil/prosil* 'sb. was extraordinarily in a hurry, sb. extraordinarily asked' do not occur at all. In the corpora of present-day texts I repeatedly came across verbs such as *(po)nравit'sja* ≈ 'to like', *(za)interesovat'* ≈ 'to interest', *vozrasti* 'to increase', *cenit'* 'to appreciate', *(vz)volnovat'* ≈ 'to agitate', *(ob)radovat'* 'to make happy', *raširjat' (sja)* 'to widen', *napominat' čto-l.* 'to resemble sth., to remind of sth.', *sposobstvovat' čemu-l.* 'to promote sth.' in the scope of *črezvyčajno*. Compare contexts (1) to (6).

- (1) Хотя и мне эта картина тоже *чрезвычайно* понравилась, но утверждение Арсения Александровича показалось несколько преувеличенным. [RNC: Михаил Козаков. Актерская книга (1978–1995)]
'Though I liked this picture very much too, the statement of Arsenij Aleksandrovič seemed rather exaggerated to me.'
- (2) – Меня *чрезвычайно* заинтересовало ваше открытие, — произнёс Красноперов. [RNC: Сергей Довлатов. Иная жизнь (1984)]
'“Your discovery has interested me very much”, Krasnoperov said.'
- (3) За последние десятилетия в самых широких кругах многих стран мира *чрезвычайно* возрос интерес к религиозному, общественному и общекультурному движению XIV века в Византии и других странах Восточной Европы. [RNC: Иоанн Мейендорф. Св. Григорий Палама, его место в предании Церкви и в современном Богословии (1992)]

‘During the last decades, the public interest in the religious, social and cultural movement in the 14th century in Byzantium and other countries of Eastern Europe grew greatly all over the world.’

- (4) Сергей *чрезвычайно* ценил это качество. [RNC: Александр Генис. Довлатов и окрестности (1998)]
‘Sergej appreciated this trait very much.’
- (5) Хэлла *чрезвычайно* взволновало это сообщение. [RNC: Валентин Бережков. Рядом со Сталиным (1998)]
‘This message extremely agitated Hell.’
- (6) Обе новости *чрезвычайно* обрадовали Гавриила [. . .]. [RNC: Борис Васильев. Были и небыли. Книга 1 (1988)]
‘Both messages made Gavriil very happy.’

To find which verbs are currently possible as elements of the construction [*črezvyčajno* P], I compiled a sub-corpus containing all RNC-texts for the period from 2000 to 2009. The total number of occurrences of *črezvyčajno* in this sub-corpus is 2 394. This degree modifier co-occurs with verbs only 126 times, which proves my thesis that it is quite atypical of *črezvyčajno* to combine with verbs. Table 1 shows with which verbs this degree adverb co-occurs in the analyzed sub-corpus.

It is easier and more economical to list the verbs which co-occur with *črezvyčajno* more or less regularly than those which do not. From the viewpoint of its lexicographic representation in terms of the Meaning↔Text Theory, it means that *črezvyčajno* has to be indicated in the value of Magn in the corresponding entries.

The restrictions in the co-occurrence of *črezvyčajno* with verbs that allow for degree modifying in principle appear not to be motivated by a given verb’s semantic field (or “thematic class” in the sense of Padučeva 2003, 2004: 30–50). Compare, for instance, the class of mental and emotional verbs. Some allow for co-occurrence with *črezvyčajno* without any restrictions (*ego éto črezvyčajno zainteresovalo* ‘all this interested him extraordinarily’, *emu éto črezvyčajno ponravilos* ‘he extraordinarily liked it’), whereas others do not (*??on črezvyčajno mečtal ob étom* ‘he extraordinarily dreamed of it’ or *??on črezvyčajno vosxiščalsja étim* ‘he extraordinarily admired it’). However, this does not mean that such restrictions are unmotivated. Certain semantic features turn out to be relevant for the combinatorial behaviour of the word *črezvyčajno*. In most cases, the features promoting the co-occurrence of verbs with this degree modifier operate only within certain semantic fields. For example, *črezvyčajno* avoids the verbs that profile the semantic feature “action” (for more detail see Dobrovol’skij [2005a]). Relatively rare exceptions include some verbs denoting abstract actions (cf. this notion in Padučeva [2004: 41]), such as

Verbs that co-occur with <i>črezvyčajno</i>	Hits (N = 126)
<i>(po)nnavit'sja</i> ≈ 'to like'	18
<i>(za)interesovat'(sja)</i> ≈ 'to interest / to get interested'	7
<i>vezēt</i> ≈ 'to be lucky', <i>(vz)volnovat'</i> ≈ 'to agitate'	5
<i>cenit'(sja)</i> 'to appreciate / to get appreciated', <i>dorožit'</i> 'to value', <i>gordit'sja čem-l.</i> 'to be proud of sth.', <i>zatrudnjat' čto-l.</i> 'to impede sth.'	4
<i>ljubit'</i> 'to love', <i>napominat' čto-l.</i> 'to resemble sth., to remind of sth.', <i>(ob)</i> <i>radovat'</i> 'to make happy', <i>osložnjat' čto-l.</i> 'to complicate sth.', <i>uvažat'</i> 'to respect', <i>vozrasti</i> 'to increase'	3
<i>nuždat'sja</i> 'to need', <i>(o)zabotit'sja</i> ≈ 'to be concerned', <i>podxodit'</i> ≈ 'to suit', <i>rasprostranit'sja</i> 'to spread', <i>rašširjat'(sja)</i> 'to widen', <i>udivljat'</i> 'to surprise, to amaze', <i>xotet'</i> 'to want', <i>zanimat'</i> 'to preoccupy', <i>zaviset'</i> 'to depend'	2
<i>bespokoit'</i> ≈ 'to disquiet', <i>imponirovat'</i> 'to make an impression', <i>l'stit'</i> 'to flatter', <i>nakalit'sja</i> 'to become strained/tense', <i>obidet'</i> 'to offend', <i>objazyvat'</i> 'to oblige', <i>oblegčat'</i> 'to facilitate', <i>obostrit'</i> 'to sharpen', <i>ogorčit'</i> 'to upset', <i>opasat'sja</i> ≈ 'to fear', <i>operežat'</i> ≈ 'to leave behind', <i>pooščrjat'</i> ≈ 'to encourage', <i>(po)vlijat'</i> 'to influence', <i>poxudet'</i> ≈ 'to grow thin', <i>preuspet'</i> 'to succeed', <i>priv-</i> <i>lekat'</i> 'to attract', <i>rasterjat'sja</i> ≈ 'to lose one's head', <i>rastvorjat'</i> 'to dissolve', <i>razdražat'</i> 'to annoy', <i>razrastis'</i> 'to expand', <i>simpatizirovat'</i> 'to sympathize', <i>smuščat'</i> 'to confuse', <i>sootvetstvovat'</i> ≈ 'to correspond', <i>sožaleť</i> 'to regret', <i>suzit'</i> 'to narrow', <i>tomit'</i> 'to torment', <i>trevožit'sja</i> 'to be anxious', <i>ukrepit'</i> 'to strengthen', <i>ukrasit'</i> ≈ 'to adorn', <i>uskorit'</i> 'to accelerate', <i>ustat'</i> 'to get tired', <i>ustraivat'</i> ≈ 'to be convenient', <i>veselit'</i> ≈ 'to cheer up', <i>vyigryvat'</i> ≈ 'to profit', <i>vyrasti</i> 'to grow', <i>zabavlat'</i> ≈ 'to amuse', <i>zagoret'sja</i> ≈ 'to get enthusiastic', <i>zatjanut'sja</i> ≈ 'to be delayed', <i>zaxvatit'</i> ≈ 'to fascinate'	1

Table 1: Co-occurrences of *črezvyčajno* with verbs in the 2000–2009 sub-corpus from RNC.

osložnjat' 'to complicate'. The more prominent the feature 'action' is in the semantic structure of a given verb, the stronger this verb is then biased against the co-occurrence with *črezvyčajno*; cf. **črezvyčajno vykatyvat' glaza* literally 'to extraordinarily roll out the eyes' or **črezvyčajno raskryť rot* 'to extraordinarily open the mouth', where the semantic feature 'action' seems to be more prominent than, for example, in *črezvyčajno osložnjat'* 'to extraordinarily complicate'.

Following this line of reasoning, we can assume that this semantic factor determines the tendency towards avoiding verbs in the scope of this modifier in general as verbs are prototypical designations of actions. Even verbs belonging to non-agentive taxonomic classes are perceived as names of certain "quasi-actions", at least to a certain degree. This is especially obvious while contrasting finite verbs with the corresponding participles or adjectives. Compare, for example,

črezvyčajno rasprostranit’/rasprostranit’sja ‘to extraordinarily spread (sth.)’ vs. *črezvyčajno rasprostraněn/rasprostraněnnij* ‘to be extraordinarily spread’. In the text corpora used here I did not encounter any instances of the co-occurrence of the former type whereas the co-occurrence of the latter type is richly presented. In the above mentioned 500 hits sample, the co-occurrence *črezvyčajno rasprostraněn/rasprostraněnnij* ‘to be extraordinarily spread’ occurs 4 times. Compare (7).

- (7) *Чрезвычайно распространенное мнение о том, что рынок освобождает, было в свое время в России важнейшем средством легитимации рыночных отношений.* [RNC: Андрей Ашкеров. Начала труда и конец производства // Отечественные записки, 2003]
 ‘The widespread opinion that the market makes free was in due course the most important means of the legitimation of exchange relations.’

Hence, Principle 1, which is based on part-of-speech categories, gains a semantic explanation. This principle seems to be the strictest one and is labelled here the Main Principle.

2.2 Semantic selection constraints

Let us briefly consider Principle 2 and Principle 3, which describe the current co-occurrence norms of the adverb *črezvyčajno*. Here we are dealing with tendencies according to which *črezvyčajno* combines (a) with words denoting non-physical properties, and (b) with words denoting features which are biased towards the positive pole of various scales, for example, with words whose semantic structure contains the feature of positive evaluation or desirability.

The relevance of feature (a) becomes particularly obvious when comparing two meanings of the same word, one of which denotes a phenomenon of the “visible world” and the second one a phenomenon of the “invisible world”. For instance, the word *prostoj* ‘simple’ forms a word combination with *črezvyčajno*, which is perceived as not quite conforming to usage from the perspective of current norms if the object being described is a physical one: a building, a piece of furniture or a dress; cf. (8).

- (8) [...] улыбаясь, сказала она Анне, которая в третьем, опять в *чрезвычайно простом*, платье вышла к ней. [RNC: Л.Н. Толстой. Анна Каренина]
 ‘[...] she said with a smile to Anna, who came in to her in a third dress, again of extreme simplicity.’

But at the same time, the lexical co-occurrence *črezvyčajno prostoj* is considered to be normative in contexts dealing with tasks, principles or situational circumstances. This is not a single case. The co-occurrence *črezvyčajno vysokij/vysoko* ‘extraordinarily high/tall/highly’ is perceived as corresponding to the current usage norms in contexts such as *črezvyčajno vysoko cenit' kogo-l./čto-l.*, *črezvyčajno vysoko otzyvat'sja o kom-l./čem-l.* ‘to estimate sb./sth. extraordinarily highly, to give high praise to sb./sth.’, *črezvyčajno vysokij uroven' bezraboticy* ‘an extraordinarily high level of unemployment’, *črezvyčajno vysokaja stoimost'* ‘an extraordinarily high cost’, *črezvyčajno vysokij rejting/status* ‘an extraordinarily high rating/status’, but not in the contexts such as *??črezvyčajno vysokij čelovek* ‘an extraordinarily tall person’ (no hits in RNC).

As far as the bias of the combinatorial profile of *črezvyčajno* towards the positive pole is concerned, it has to be stressed that here we are dealing not only with the opposition of ‘good’ or ‘desirable’, on the one hand, and ‘bad’ or ‘undesirable’, on the other,⁵ but also with all other semantic scales, whose poles can be interpreted as positive vs. negative; compare the contrasting of ‘big’ and ‘small’, of ‘long/tall’ and ‘short’, and the like. For example, the co-occurrence *črezvyčajno vysokij* ‘extraordinarily high’ is found in the Russian National Corpus 120 times, whereas *črezvyčajno nizkij* ‘extraordinarily low’ only 54 times. In such dichotomies it is not always clear which element is positive and which is negative if the situation as a whole is taken into account;

“Positive Pole” co-occurrences	RNC hits	“Negative Pole” co-occurrences	RNC hits
<i>črezvyčajno vysoko</i> ‘extraordinarily highly’	38	<i>črezvyčajno nizko</i> ‘extraordinarily low’	9
<i>črezvyčajno aktivno</i> ‘extraordinarily actively’	6	<i>črezvyčajno passivno</i> ‘extraordinarily passively’	0
<i>črezvyčajno jarko</i> ‘extraordinarily bright’	14	<i>črezvyčajno bledno</i> ‘extraordinarily dimly’	2
		<i>črezvyčajno tusklo</i> ‘extraordinarily wanly’	0

Table 2. Frequency of polarity co-occurrences from RNC

⁵ The sensitivity of the word *črezvyčajno* to evaluative properties of its contextual partners was pointed out in previous research; cf., for example, Červenková (1975: 17).

think of pollution and oxygen levels in the atmosphere—low and high, respectively, would be close to the positive pole in the sense of ‘good’. However, what is really important for relevant combinatorial preferences is the isolated meaning of a given adjective, rather than the situational sense of the NP. That is to say, the combinatorial properties of *črežvyčajno* are not sensitive to the [+/-] feature that can be ascribed to the situation in question in the sense of dichotomy of ‘good’ and ‘bad’, but they are sensitive to the [+/-] feature fixed in the meaning of the immediate co-occurrence partner (if a scalar interpretation can be applied to its semantics).

Of course, this tendency should not be taken as a strict rule. It expresses itself in statistical preferences rather than in strict constraints (see Section 3). Co-occurrences such as *črežvyčajno vysoko/aktivno/jarko* ‘extraordinarily highly/actively/bright’ are encountered in present-day texts much more often than co-occurrences such as *črežvyčajno nizko/passivno/bledno/tusklo* ‘extraordinarily low/passively/dimly/wanly’ (*bledno* ‘dimly’ and *tusklo* ‘wanly’ can both be considered antonyms of *jarko* ‘bright’). Compare Table 2.

This holds not only for Principle 3, but also for Principle 2. But counterexamples to both principles, which sound quite normative, can be found in corpora. Compare *črežvyčajno poxudet’* ≈ ‘to grow extraordinarily thin’ and *črežvyčajno ustat’* ‘to get extraordinarily tired’ from Table 1, as well as context (9) where *črežvyčajno* combines with a word denoting physical property (violation of Principle 2) and denoting a negative-pole feature (violation of Principle 3).

- (9) Перед судебным следователем стоит маленький, *чрезвычайно тощий* мужичонка в пестрядинной рубахе и латаных портах. [RNC: А.П. Чехов. Злоумышленник]
 ‘An exceedingly lean little peasant, in a striped hempen shirt and patched drawers, stands facing the investigating magistrate.’

Semantic selection constraints described here resemble the category of “slot restrictions” as explained in Kuiper (2009: 19): “While the syntactic category of a slot constraints what it may contain syntactically, there are frequently other constraints of an arbitrary kind”.

2.3 Genre restrictions of *črežvyčajno*

An important group of factors that govern the combinatorial properties of the adverb *črežvyčajno* must be described in terms of stylistics or discourse specifics. The most significant parameter is here the type of discourse, i.e. the set of thematic domains and stylistic registers that form a natural environment for

the word *črezvyčajno*. This seems to be the common parameter for the lexical items which, in spite of their semantic heterogeneity, can potentially occur in the scope of this modifier. There is a certain correlation between this factor and the bias towards co-occurrence with names of properties that describe non-physical entities; discussing abstract issues rather than outward, physical properties of things and people is more typical of discourse types which can be labelled “formal”. Thus, it can be assumed that the “non-physical tendency”, i.e. Principle 2 is governed by the relevant stylistic or discursive features of the adverb *črezvyčajno*.

Genre factors, namely the markedness of the word *črezvyčajno* as “non-ordinary”, manifest themselves, above all, in this word being perceived as more appropriate in contexts of “formal discourse” rather than in thematic domains typical of everyday language. Furthermore, this factor influences the stylistic parameters of contextual partners of the word *črezvyčajno*. Compare, for example, *črezvyčajno alčnyj* ≈ ‘extraordinarily acquisitive’ vs. the stylistically less acceptable co-occurrence *črezvyčajno žadnyj* ‘extraordinarily greedy’.

The fact that genre restrictions and semantic features proper are interrelated is often stressed in linguistics, especially within the theoretical framework of Construction Grammar. “The term ‘meaning’ is intended to represent all of the CONVENTIONALIZED aspects of a construction’s function, which may include not only properties of the situation described by the utterance but also properties of the discourse in which the utterance is found [. . .] and of the pragmatic situation of the interlocutors” (Croft 2001: 19).

The next two sections contain the quantitative analysis of empirical data. It is aimed at proving the validity of the principles discussed above. Here I employ two different methods: corpus data and survey data analysis. Not only is the verification of the assumptions about the principles governing the co-occurrence behaviour of the given degree modifier an issue, but also the question as to how corpus evidence relates to survey data.

3 Corpus data analysis

The data analysis undertaken in this paper is corpus-based rather than corpus-driven by nature. According to Tognini-Bonelli (2001), corpus-based approaches give priority to the pre-existing hypotheses or theoretical statements, which are validated using corpus data, whereas corpus-driven linguistics starts with empirical corpus data and their (mostly statistical) analysis, which leads to a hypothesis building.

Analysis of three random samples from the Russian National Corpus (RNC) was conducted. Sample 1 contains 100 hits from Russian newspapers in the period between 2000 and 2008. Sample 2 is based on present-day texts (both fiction and journalism) published between 2003 and 2008 (100 hits). Sample 3 consists of 100 hits from the 19th century. Every use of *črezvyčajno* in the analysed occurrences was tagged according to the three principles discussed in the previous sections. Compare, for example, contexts (10) and (11) from Sample 3.

- (10) Бульба вскочил на своего Черта, который бешено отшатнулся, почувствовав на себе двадцатипудовое бремя, потому что Тарас был *чрезвычайно тяжел и толст*. <A> <Ph> <Pos> [RNC: Н.В. Гоголь. Тарас Бульба]
 ‘Bulba sprang upon his “Devil,” which bounded wildly, on feeling on his back a load of over thirty stone, for Taras was extremely (literally ‘extraordinarily’) stout and heavy.’
- (11) Никому в свете Нежданов не говорил о своих сношениях с ним и *дорожил ими чрезвычайно*. <V> <NPh> <Pos> [RNC: И.С. Тургенев. Новь]
 ‘He had never told anyone of his relation with Silin, a relation that was very dear to him.’

The tags <A> <Ph> <Pos> in (10) mean that in *črezvyčajno tjažěl i tost* ‘extraordinarily stout and heavy’ *črezvyčajno* combines with adjectives <A> denoting a physical property <Ph> which is biased towards the positive pole <Pos> of a dimensional scale. The tags <V> <NPh> <Pos> in (11) mean that here *črezvyčajno* combines with a verb <V> denoting a non-physical attitude <NPh> with a positive evaluation <Pos>. This method of analysis has certain limitations. I evaluated the occurrences of *črezvyčajno* myself and performed the tagging. A very similar method of analysis was applied in Divjak and Gries (2006). Table 3 shows the results which were obtained.

In the first three columns of Table 1 different combinations of the discussed principles are presented. Sign [+] in the first column means that Principle 1 is followed, i.e. *črezvyčajno* combines with adjectives or adverbs, sign [–] means that *črezvyčajno* combines with verbs. In the second column, [+] means that the word in the scope of this degree modifier denotes non-physical properties, [–] refers to its physical properties. In the third column, signs [+] and [–] distinguish positive and negative poles of various scales.

The frequency distributions in the samples from the 21st century show a very high degree of similarity. The cases where all three tendencies discussed in the previous sections (Principle 1, Principle 2 and Principle 3) are followed are most frequent (68–70%). The violation of Principle 1, as well as the violation of

Adjective or Adverb <A> or <Adv>	Non-Physical Properties <NPh>	Positive Pole <Pos>	Percentage		
			21st century, Sample 1 (N=100)	21st century, Sample 2 (N=100)	19th century (N=100)
+	+	+	68	70	34
+	+	–	21	20	18
+	–	+	3	3	8
–	+	+	5	3	20
+	–	–	1	2	9
–	+	–	2	2	8
–	–	+	0	0	1
–	–	–	0	0	2

Table 3: Frequency of tagged features.

Principle 2 is possible in present-day texts, but appears quite rare. These two principles are never violated at the same time, i.e. it is impossible in present-day usage for *črezvyčajno* to combine with verbs denoting physical actions (however, physical states and changing states are possible as rare exceptions, cf. Table 1). Principle 3 seems to be not so strict, i.e. words denoting properties which are biased towards the negative pole may be modified by *črezvyčajno* (20–21%).

The usage of this degree adverb in the 19th century differs from the usage in the present in several respects. First of all, its combinatorial profile was much broader, i.e. the selection constraints were not as strict as they are now. The question why the combinatorial profile of this word has changed over the last hundred and fifty years cannot be answered along the lines of some general trends of language development, such as, e.g., grammaticalization (cf., inter alia, the results of the corpus-based diachronic study of the English intensifier *very much* by González-Díaz 2008). The reasons for this seem to be rather conventional by nature. As Croft (2000: 231–232) pointed out, “conventions themselves are never 100% invariant; conventions are always in the process of being acquired, replaced or lost”.

Occurrences in which all three principles are observed are still the most frequent (34%). In contrast to the 21st century, the violation of the Main Principle (Principle 1) in the 19th century was rather frequent (20%). The violation of Principle 1 and Principle 2 at the same time was quite rare, but still possible. The violation of selection restrictions based on Principle 3 was acceptable, almost to the same extent as it is today (18%).

These results showing linguistic changes over time confirm our assumptions about the structure of the combinatorial profile and its possible dynamics.

Typically, the combinatorial profile of a given word has a core and a periphery. The central part of the combinatorial profile, i.e. its core, remains stable whereas its periphery is responsible for linguistic change. Peripheral co-occurrences of *črezvyčajno* (in which some of its co-occurrence principles are violated) can change their status in due course, so that most of them would have been ruled out by present-day usage norms.

4 Survey data

4.1 Hypotheses

Another method that I used for the verification of the proposed principles was a survey. It operated with two parameters: Normativity and Frequency. The aim of this survey was to find out whether native speakers perceive expressions that violate the principles formulated above as odd or unusual. The respondents had to estimate the proposed expressions both in terms of perceived usage norms and in terms of their perceived occurrence frequency. This means that respondents participating in this survey functioned as experts. This explains why only well-educated native speakers (most of whom were linguists) were selected for the survey. Theoretically these two parameters do not need to coincide because native speakers may find a certain expression quite normative though it seems to have a low frequency of occurrence (cf., for instance, Wray 2002: 277). Also Arppe and Järviö (2007: 131) point to the fact that “infrequency does not categorically always entail substantially lower acceptability”. One of the tasks of the conducted survey was to test whether there is a correlation between Normativity and Frequency.

Two experimental hypotheses were put forward.

Hypothesis 1. Native speakers will estimate the Normativity and Frequency of expressions belonging to the core of the combinatorial profile higher than for expressions belonging to its periphery or ruled out by the usage norms. Co-occurrences are considered to be peripheral in terms of the combinatorial profile if they violate, at least, one of the proposed principles. The Normativity and Frequency of expressions ruled out by the norms of present-day usage will be estimated lower than these parameters of all the other expressions.

Hypothesis 2. The estimates of expressions occupying the periphery of the combinatorial profile will vary more than the estimates of expressions belonging to the core of the combinatorial profile, and more than the estimates of expressions which are completely ruled out by the usage norms. This suggestion is based on the assumption that the periphery of a combinatorial profile allows for more subjective freedom in language use than expressions that are absolutely conventional and expressions that are perceived as incorrect.

4.2 Method

Twenty-nine educated native speakers of Russian made up the group of experts. All of them had a level of education not lower than an MA degree in different fields, 26 of them being linguists. The experts were given in written form three types of lexical co-occurrences with *črezvyčajno*:

- (1) expressions which did not violate Principles 1 to 3 as well as the relevant discourse factors, i.e. the core of the combinatorial profile of this word,
- (2) expressions which did not violate Principle 1 (the Main Principle), but did not clearly follow Principle 2 and/or Principle 3, and/or did not observe the relevant discourse factors in all cases, i.e. expressions belonging to the combinatorial profile's periphery, and
- (3) expressions in which Principle 1, as well at least one of the remaining principles – Principle 2 or Principle 3 – and/or the relevant genre and register conditions were violated, i.e. expressions which are expected to be ruled out by the norms of present-day usage.

First, the experts estimated the compliance of these expressions with the lexical norms (Normativity) using the 5-point scale:

- 1 = not possible
- 2 = hardly possible
- 3 = possible, but sounds odd
- 4 = conforms to usage norms, but still somewhat strange
- 5 = absolutely conforms to usage norms

Then the experts estimated the subjective frequency (Frequency) of these expressions using another 5-point scale:

- 1 = never
- 2 = very rarely
- 3 = rarely, rather than often
- 4 = often, rather than rarely
- 5 = often

The experts gave their answers in written form.

The following 20 phrases with *črezvyčajno* were given to the experts:

1. *чрезвычайно успешно* (*črezvyčajno uspešno*) 'extraordinarily successful'
2. *чрезвычайно скоро* (*črezvyčajno skoro*) 'extraordinarily fast/quick'
3. *чрезвычайно важный секрет* (*črezvyčajno važnyj sekret*) 'an extraordinarily important secret'

4. *чрезвычайно расхохотался* (*črezvyčajno rasxoxotal'sja*) 'extraordinarily burst out laughing'
5. *чрезвычайно простой вопрос/ответ* (*črezvyčajno prostoj vopros/otvet*) 'an extraordinarily simple question/answer'
6. *чрезвычайно интересные вещи* (*črezvyčajno interesnye vešč'i*) 'extraordinarily interesting issues'
7. *чрезвычайно доверчив* (*črezvyčajno doverčiv*) 'extraordinarily trustful'
8. *чрезвычайно сложные звуки* (*črezvyčajno složnye zvuki*) 'extraordinarily complex sounds'
9. *чрезвычайно неустойчивый* (*črezvyčajno neustojimyj*) 'extraordinarily tireless'
10. *чрезвычайно стройная фигура* (*črezvyčajno strojnaja figura*) 'an extraordinarily slender body'
11. *чрезвычайно страшный* (*črezvyčajno strašnyj*) 'extraordinarily scary/fearful'
12. *чрезвычайно ошиблась* (*črezvyčajno ošiblas'*) '[she] was extraordinarily mistaken' (In Russian [degree adverb_{Magn} V_{past}])
13. *бумажки чрезвычайно испортились* (*bumažki črezvyčajno poportilis'*) 'papers were extraordinarily spoiled' (In Russian [N degree adverb_{Magn} V_{past}])
14. *чрезвычайно сильное воображение* (*črezvyčajno sil'noe voobraženie*) 'extraordinarily strong imagination'
15. *чрезвычайно любопытная болезнь* (*črezvyčajno ljubopytnaja bolez'n*) 'an extraordinarily interesting disease'
16. *чрезвычайно вкусный обед* (*črezvyčajno vkusnyj obed*) 'an extraordinarily delicious dinner'
17. *держаться чрезвычайно прямо* (*deržat'sja črezvyčajno prjamo*) 'to hold oneself extraordinarily upright'
18. *чрезвычайно весело* (*črezvyčajno veselo*) 'extraordinarily cheerfully'
19. *чрезвычайно старый и грязный* (*črezvyčajno staryj i grjaznyj*) 'extraordinarily old and dirty'
20. *одетая чрезвычайно к лицу* (*odetaja črezvyčajno k licu*) 'dressed extraordinarily suitable'

Expressions 1, 3, 5, 6, 7, 14 belong to Group 1, expressions 4, 12, 13 to Group 3, and the rest of co-occurrences forms the periphery of the combinatorial profile of *črezvyčajno* and, therefore, belongs to Group 2. Expression 9 was excluded from further analysis because it revealed a certain ambiguity since it could be understood both in a physical and in a non-physical sense.

Due to logistical reasons I could not include an equal number of expressions in Groups 1, 2 and 3. Nevertheless, the structure of obtained data allows for statistical analysis and further interpretation.

4.3 Results

Standard statistical procedures were applied for the data analysis.

Consistency of experts' estimates was very high (Cronbach's alpha 0.97 both for Normativity and for Frequency). This allows us to conduct further analysis.

Mean scores for the three groups are presented in Table 4.

The Kruskal-Wallis test revealed statistical significance of differences among means in these three groups ($\chi^2 = 12.174$, $p = 0.002$ for Normativity, $\chi^2 = 12.481$, $p = 0.002$ for Frequency). Paired comparison of means showed that there were statistically significant differences between groups 1 and 2 (Mann-Whitney's $U = 3$, $p = 0.001$ for Normativity, Mann-Whitney's $U = 4$, $p = 0.002$ for Frequency) and between groups 2 and 3 (Mann-Whitney's $U = 3$, $p = 0.025$ for Normativity, Mann-Whitney's $U = 1$, $p = 0.007$ for Frequency).

The results obtained confirm the experimental Hypothesis 1 and show that the expressions which follow the proposed principles receive higher scores than expressions which violate them.

In order to test the experimental Hypothesis 2, the ranges of the experts' estimates in each group were calculated (Table 5).

We can see from Table 5 that the scores were most scattered in Group 2, which is consistent with the previous suggestion that the estimates of the expressions of Group 2 would highly depend on the individual preferences of speakers.

All the results presented above show that both variables Normativity and Frequency behave almost identically. The Spearman correlation coefficient between

	Group 1	Group 2	Group 3
Normativity	3.96	2.89	1.30
Frequency	3.52	2.52	1.15

Table 4: Mean scores of experts' estimates.

	Group 1	Group 2	Group 3
Normativity	0.93	2.48	0.24
Frequency	1.21	2.07	0.29

Table 5: Ranges of experts' estimates

them is 0.97 ($p < 0.001$). This means that both survey parameters were practically identical for experts. For this reason, it is sufficient to use only one of them in similar research.

5 Discussion

5.1 Corpus evidence vs. survey data

The results obtained show that, in general, the survey data confirm the assumptions made on the basis of corpus evidence. However, these results cannot be regarded as definite.

For one thing, recent research provides different findings. An analysis of some adjectival and adverb degree modifiers of Russian provided by Tixonovič (2009) showed a relatively significant divergence between corpus-based and experimental data. However, the starting point of her analysis is a sample of lexical items which can be modified by various intensifiers, rather than a concrete degree modifier and its combinatorial profile, as is the case in my paper. Therefore, it is not surprising that the analysis by Tixonovič shows different results. Since the choice of an intensifier in a given constructional setting is not strictly bound, the speakers have much more freedom in their individual decisions as to which specific degree modifier to use in every context. The question which I am trying to answer here is whether there are certain constraints on the use of a single intensifier, what these constraints are, and how they can be explained. The tendencies which govern the combinatorial preferences of a genre-restricted word such as *črezvyšajno* look much more like the basic principles of its co-occurrence than the tendencies predicting the choice of an intensifier in combination with certain lexemes.

There is also another reason why my results may not be considered to be definitive. The results obtained have some limitations. First, only one word was analysed in its contextual behaviour. Second, it was myself who decided to what extent the proposed three principles, as well as relevant discourse factors, were violated in each of the expressions used as stimuli. It would have been better if these conditions were estimated by independent experts. Third, the variation of principle violations could have been more systematic, so that the violation of every principle varied independently of other deviations. In addition, the experimental design can be improved upon. My experiment presented phrases rather than larger contexts (whole sentences). The experts received items in the same order with no distractors, and they were a rather homogeneous group of well-educated native speakers. For these reasons, further studies should be conducted in this domain in order to obtain more definitive results.

5.2 Towards a linguistic explanation of the observed selection restrictions

An interpretation of the results must attempt to find some common properties that are shared by the combinatorial principles discussed in the previous sections. Though these principles appear to be independent of each other, at a deeper level, they may reveal common features that can explain more general phenomena. Perhaps general properties in the semantics and/or stylistics of the words in question can be found, which could explain the observed selection constraints and preferences. As already mentioned in Section 1, selection constraints of degree adverbs do not necessarily need to be traced back to the semantic structure of these adverbs. Sometimes we are dealing with what are called here usage-based combinatorial restrictions or preferences. However, it is instructive to look for some possible correlations between the semantic structure of *črežvyčajno* and its co-occurrence behaviour. For this reason, we need to find at least a tentative definition of the degree modifier *črežvyčajno*, or, to be more precise, of the construction [*črežvyčajno* P]:

črežvyčajno P ≈ 'P in a very high degree; the speaker claims that the degree of P is so high that it exceeds the bounds of the domain of ordinariness'

I understand “the bounds of the domain of ordinariness” as referring to the usual bounds of variability of the feature P, i.e. such bounds of its variability within which the degree of P is perceived as not violating habitual expectations. This domain includes also a high degree of P, that is, the meaning ‘very P’.

Let us compare this definition with definitions of near-synonymous degree modifiers, such as *beskonečno*, *bezgranično*, *bezmerno* ≈ ‘infinitely’—in Grigor’eva (2004)⁶—or *krajne*, which can be approximately defined in the following way:

krajne P ≈ ‘P in a very high degree; the speaker claims that the degree of P is so high that it is near P’s highest possible bound’

The relevant semantic differences between these degree modifiers evidently influence the meaning of the whole phrase containing them: *krajne glupyj čelovek* literally ‘an utterly stupid man’ is not quite the same as *beskonečno glupyj čelovek*

⁶ Cf.: “*бесконечно <безгранично, безмерно> P* = ‘P в очень высокой степени; говорящий считает, что степень P настолько высока, что это трудно представить’” [‘P in a very high degree; the speaker claims that the degree of P is so high that it is difficult to imagine’] (Grigor’eva 2004: 15).

literally ‘an infinitely stupid man’. These semantic differences can be traced back to the specifics of the etymological basis or “the inner form” (of German and Russian linguistic tradition), which may express itself in the actual meaning of lexical items. For more detail on the nature of the inner form see Dobrovol’skij (1996); Baranov and Dobrovol’skij (1998, 2008); Zaliznjak (1998). Semantic differences of this kind turn out to be relevant especially in contexts profiling just those semantic components that are inherited from the inner form. Nevertheless, since the traces of the source concept are present in latent form in the semantics of a given lexical item, it can be assumed that they can have an impact on the combinatorial profile.

This basic assumption lays the foundation for a hypothesis about certain combinatorial properties resulting from the inner form of the adverb *črezvyčajno*. From the definition of this word, the assumption is that its preferred co-occurrence demands lexical items denoting properties which contain the idea of the DOMAIN OF ORDINARINESS. The idea of the DOMAIN OF ORDINARINESS is related to the notion of NORM (if ordinariness is viewed objectively, which is the case here). In the present paper I prefer not to use the word *norm* as an element of metalanguage because of its polysemy. Compare similar ideas in Uryson (2000). Therefore, with regard to the DOMAIN OF ORDINARINESS, it should be explicitly stressed that here we are dealing with the standard norm (i.e. with the idea of the characteristics typical of a given situation or a given object) rather than with the deontic or teleological norm.

It can be assumed that the feature ‘domain of ordinariness’ is, to a certain extent, responsible for the combinatorial properties of the word *črezvyčajno*. By virtue of the relatively transparent morphological structure of this word, its inner form addresses the idea of bounds of habitual expectations, rather than the NORM as a point on a certain scale. Connections of the adverb *črezvyčajno* with the idea of exceeding the bounds of ordinariness are also maintained by the semantics of the adjective *črezvyčajnyj*, not primarily in the corresponding meaning (cf. *črezvyčajno udivlěn—črezvyčajnoe udivlenie* ‘extraordinarily astonished—extraordinary astonishment’), but in the meaning realised in co-occurrences, such as *črezvyčajnyj s”ezd* ‘extraordinary congress’, *črezvyčajnye mery* ‘extraordinary measures’.

Following this line of reasoning, we can hypothesize that the feature ‘domain of ordinariness’ forms a semantic bridge between some of those seemingly isolated combinatorial principles discussed above. It is extremely difficult to postulate the DOMAIN OF ORDINARINESS for actions such as “staring” or “laughter”. Such actions hardly presuppose some habitual expectations. Obviously, this is the reason why the co-occurrences *črezvyčajno vykatyvala glaza* literally ‘[she] extraordinarily rolled out her eyes’ and *črezvyčajno rasxotalsja* ‘extraordinarily

burst out laughing' used by Dostoevskij are perceived as norm violating. Similar occurrences could hardly be found in texts from other authors of the 19th century (see Section 3). Compare, however, context (12).

- (12) Паклин *чрезвычайно* рассмешил старичков известным гоголевским анекдотом [. . .]. [RNC: И.С. Тургенев. Новь]
'Paklin amused them very much by relating the well known Gogol anecdote [. . .].'

Co-occurrences of this kind were perceived as being marginal even then. They belonged to the extreme periphery of the combinatorial profile of *črezvyčajno*, and were ruled out by the usage norm of subsequent decades. The reason for this is that such co-occurrences contradict the inner form of this adverb.

Nevertheless, in general it is difficult to relate the bias of *črezvyčajno* towards co-occurrence with words that lack the feature 'action' to the idea of the domain of ordinariness. Based on intuition, we can assume that "states" and "properties", as a whole, associate with the DOMAIN OF ORDINARINESS in a more natural way than 'actions'. 'Actions' are associated with a point on a timescale, rather than with a domain of any kind. 'States' and 'properties', however, are thought of as being relatively long-continued and stable, so that, metaphorically, they can be put into a kind of domain more easily. In general, it seems to be the case that the bias of the degree modifier *črezvyčajno* against co-occurrence with verbs (Principle 1) is partially motivated by its meaning. But such hypotheses need to be backed up by independent empirical data. Moreover, it is evident that the bias towards the positive pole (Principle 3) correlates to some extent with the idea of the domain of ordinariness, because the negative values of various parameters are perceived, in the naive model of the world, as norm violation, and therefore, do not easily correlate with the idea of habitual expectations. Admittedly, such an interpretation of the concept of ORDINARINESS relating back to the idea of the standard norm receives an additional "deontic touch". As for the semantic feature 'non-physical', typical of *črezvyčajno*'s contextual partners (Principle 2), it can hardly be connected with the idea of ORDINARINESS. In this case, factors of discourse type seem to have more explanatory power (cf. Section 2.3) because, in "formal discourse", the issues discussed are mostly abstract, non-physical entities.

It is these stylistic and discursive properties that distinguish the modifier *črezvyčajno* from its nearest synonyms *neobyčajno* and *neobykovenno*. Being not bound to a special type of discourse, the latter two words do not display any selection constraints concerning lexical items denoting physical properties. For example, *neobykovenno xoroša soboj* 'exceptionally good-looking' sounds better than *črezvyčajno xoroša soboj* 'extraordinarily good-looking', and *neobyčajno bol'sie glaza* 'exceptionally big eyes' is better than *črezvyčajno bol'sie glaza*

‘extraordinarily big eyes’.⁷ Thus with regard to the combinatorial profile, register phenomena and factors of discourse type are not less important than semantics.

The general impression is that it is not possible to trace all of the discussed combinatorial tendencies of the adverb *črezvyčajno* back to its semantic structure and/or its genre properties. In these cases we are rather dealing with some weak correlations. This means that some lexical co-occurrences of the form [*črezvyčajno* P] have to be regarded as word combinations that are neither really free nor phraseologically bound, in the strict sense of the word.

6 Conclusion

The aim of the present paper is theoretical in nature; it attempts to discover some significant features of lexical co-occurrence. The linguistic focus of my investigation is on the Russian degree modifier *črezvyčajno*, which can be literally translated as ‘extraordinarily’. It displays some combinatorial restrictions, which were discovered here and generalized as three basic principles. The analysis of these principles showed that factors motivating the combinatorial properties of a lexical item can be either of a **semantic** or **discursive** nature, or its co-occurrence can be driven by combinatorial preferences which can be traced back neither to semantics nor to discourse phenomena. Such properties were labelled here **usage-based restrictions of preferences**.

It is a well-known fact that the co-occurrence of degree modifiers (as well as words of many other semantic classes) is lexicalised to a large extent. For this reason, the existence of usage-based preferences, which are irregular and unpredictable, is naturally determined. Nevertheless, the combinatorial behaviour of the word *črezvyčajno* turns out to be motivated to a certain degree. It is reasonable to postulate a sphere of co-occurrence which takes an intermediate position between semantically constrained and purely lexically constrained co-occurrence—in the sense of Apresjan (1974). The selection of potential contextual partners is neither completely arbitrary (as it is the case with the lexical co-occurrence) nor strictly predictable (as it is the case with the semantic co-occurrence). This “intermediate” type of co-occurrence has to be described in terms of tendencies or principles, which mainly operate within certain lexical classes and can be investigated within both the theoretical framework of Construction Grammar and the Meaning↔Text Theory.

⁷ There are 6 hits with *neobyknovenno xoroša soboj* in the Russian National Corpus, plus 9091 Yandex-hits and only 3 hits in the RNC plus 1041 Yandex-hits for *črezvyčajno xoroša soboj*. There are 318 Yandex-hits for *neobyčajno bol'shie glaza*, and only 9 for *črezvyčajno bol'shie glaza*.

In contrast to traditional linguistic theories, Construction Grammar does not accept the strict borderline between free and bound phenomena of language. In other words, according to Construction Grammar, there is no clear opposition of co-occurrences that are, on the one hand, fully compositional and rule-governed and, on the other hand, unpredictable phenomena of phraseology. The data discussed in the present paper confirms these ideas. Although no formal apparatus for a precise and elegant constructional representation of such “intermediate” phenomena has been developed so far (at least, to my knowledge), the theoretical impetus of Construction Grammar is very stimulating for investigating co-occurrences of this kind as well as searching for appropriate formats of their lexicographic description. As for finding out appropriate metalinguistic tools for representing such combinatorial phenomena, the apparatus of Lexical Functions seems to be the most useful instrument at the moment. The postulate that a given Lexical Function can be linguistically expressed quite differently depending on the keyword (the base of the collocation) explains why *črezvyčajno* cannot be altered by other degree adverbs with the semantics of Magn in all cases of the construction [*črezvyčajno* P]. However, the metalanguage of Lexical Functions works much better in “classical” cases of restricted collocations, i.e. in cases in which all given co-occurrences can be listed exhaustively as phraseologically bound items.

From the lexicographic perspective, my research shows that combinatorial properties of this kind have to be specified in the dictionary (by pointing to the relevant tendencies and principles of co-occurrence as well as by exemplifying them by most typical and frequent word combinations from text corpora). This solution seems to be most adequate because it is impossible to give a reliable semantic or genre-based rule that would license all usage-conforming instances of [*črezvyčajno* P] and exclude all instances that contradict the norms of present-day usage. Cf. similar argumentation in (Iordanskaja and Mel'čuk 2011) concerning the ability of some Russian verbs to participate in parenthetical constructions.

Corpus evidence and survey data are used in this paper as additional tools to validate my hypothetical statements. Nevertheless, the relationship between these two sources of linguistic evidence is the second focus of this paper. The cognitive reality of corpus data needs empirical verification. A priori it cannot be claimed that generalizations about language use obtained on the basis of text corpora (even if they are large and representative enough) reflect the linguistic competence of native speakers; cf., e.g. (Bybee and Hopper 2001; Nordquist 2004; McGee 2009). Wray (2002: 277) also points to the fact that “patterns of knowledge” (i.e. linguistic intuition) cannot be equated with “patterns of use” (i.e. corpus evidence). Theoretically it is possible that these two data sources exist independently of each other, to a certain extent. Compare, for example, Gilquin

and Gries (2009: 17) who stress that “it is still true that the relation between the two types of data remains unclear and that identity cannot be taken for granted”. It is also possible that correlations between corpus evidence and survey data turn out to be weak. Here, an attempt to test the correlation between corpus evidence and estimations of the corresponding expressions by native speakers has been undertaken.

Analysis based only on testing corpus data and not including people’s intuition is not sufficient as an investigation method. Instead, it is necessary to find ways to combine various methods (based on both corpora and surveys) in order to investigate the semantics, stylistics and usage-based behaviour of linguistic units.⁸ The results of my investigation combining these two methods of analysis show a strong convergence of corpus-based and survey data. Though these results are insufficient to be regarded as definite, as a whole, it can be concluded that the survey data confirm the suggestions made on the basis of corpus evidence.

Russian Academy of Sciences, Russian Federation

References

- Apresjan, Jurij D. 1974. *Leksičeskaja semantika: Sinonimičeskie sredstva jazyka* [Lexical semantics: Synonymous means of language]. Moskva: Nauka.
- Apresjan, Jurij D. 2004. O semantičeskoj nepustote i motivirovannosti glagol’nyx leksičeskix funkcij [Verbal Lexical Functions are motivated and not empty semantically]. *Voprosy jazykoznanija* 4. 3–18.
- Apresjan, Jurij D. 2008. O proekte aktivnogo slovarja (AS) russkogo jazyka [Project of an active dictionary of Russian]. In *Dialogue’08: Computational linguistics and its applications*, 23–31. Moskva: Rossijskij gosudarstvennyj gumanitarnyj universitet.
- Arppe, Antti & Juhani Järviö. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131–159.
- Baranov, Anatolij N. & Dmitrij O. Dobrovol’skij. 1998. Vnutrennjaja forma idiom i problema tolkovanija [Inner form of idioms and the problem of definition]. *Izvestija AN, Serija literatury i jazyka* 57(1). 36–44.
- Baranov, Anatolij N. & Dmitrij O. Dobrovol’skij. 2008. *Aspekty teorii frazeologii* [Aspects of the theory of phraseology]. Moskva: Znak.
- Bybee, Joan & Paul Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 1–24. Amsterdam: John Benjamins.

⁸ Compare the Special Issue of Corpus Linguistics and Linguistic Theory 5 (1) 2009 “Corpora and experimental methods”, edited by Gaëtanelle Gilquin. Compare also (Nordquist 2004; Gries et al. 2005; Kepser and Reis 2005; Arppe and Järviö 2007; Mollin 2009; Tixonovič 2009; Divjak and Gries to appear).

- Červenková, Irina V. 1975. *Obščie adverbial'nye pokazateli mery priznaka (v sovremenom ruskom literaturnom jazyke)* [Adverbial degree modifiers of general semantics (in present-day literary Russian)]. Moskva: Institut russkogo jazyka AN SSSR dissertation.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow: Longman.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23–60.
- Divjak, Dagmar & Stefan Th. Gries (eds.). to appear. *Corpus and cognition: Converging and diverging evidence*. Berlin & New York: Mouton de Gruyter.
- Dobrovol'skij, Dmitrij O. 1996. Obraznaja sostavljajuščaja v semantike idiom [The image component of the idiom semantics]. *Voprosy jazykoznanija* 1. 71–93.
- Dobrovol'skij, Dmitrij O. 2005a. Faktory sočetaemosti: Semantika, pragmatika, uzus [Factors of combinatorics: semantics, pragmatics, usage]. *Russkij jazyk v naučnom osveščanii* 2. 43–86.
- Dobrovol'skij, Dmitrij O. 2005b. Semantic motivation of Lexical Functions and its limits (The combinatorial profile of the Russian degree modifier *črezvyščajno*). In Jurij D. Apresjan & Leonid L. Iomdin (eds.), *East West encounter: Second international conference on Meaning-Text Theory*, 110–121. Moscow: Slavic Culture Languages Publishing House.
- Fillmore, Charles J. 2008. Border conflicts: FrameNet meets Construction Grammar. In Elisenda Bernal & Janet DeCesaris (eds.), *Proceedings of the XIII EURALEX international congress*, 49–68. Barcelona: IULA.
- Gilquin, Gaëtanelle & Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago & London: The University of Chicago Press.
- González-Díaz, Victorina. 2008. Recent developments in English intensifiers: The case of *very much*. *English Language and Linguistics* 12(2). 221–243.
- Gries, Stefan Th. 2002. Evidence in linguistics: Three approaches to genitives in English. In Ruth M. Brend, William J. Sullivan & Arle R. Lommel (eds.), *LACUS Forum XXVIII: What constitutes evidence in linguistics?*, 17–31. Fullerton: LACUS.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4). 635–676.
- Grigor'eva, Svetlana A. 2004. Slovarnaja stat'ja *beskonečno* . . . [The dictionary entry *endless* . . .]. In Jurij D. Apresjan (ed.), *Novyj ob"jasnitel'nyj slovar' sinonimov russkogo jazyka* [The New dictionary of Russian synonyms]. 2nd edn., 15–17. Moskva & Wien: Jazyki slavjanskoj kul'tury & Wiener Slawistischer Almanach.
- Iordanskaja, Lidija & Igor Mel'čuk. 2011. Illocutive parenthetical verbs in Russian. In *Fifth international conference on Meaning↔Text Theory, Barcelona 2011. Proceedings*. <http://meaningtext.net/mtt2011/proceedings/papers/IordanskajaMelcuk.pdf> (accessed 15 October 2011).
- Kepser, Stephan & Marga Reis (eds.). 2005. *Linguistic evidence: Empirical, theoretical and computational perspectives*. Berlin & New York: Mouton de Gruyter.
- Kuiper, Koenraad. 2009. *Formulaic genres*. New York, Melbourne etc.: Palgrave Macmillan.
- McGee, Iain. 2009. Adjective-noun collocations in elicited and corpus data: Similarities, differences, and the whys and wherefores. *Corpus Linguistics and Linguistic Theory* 5(1). 79–103.

- Mel'čuk, Igor A. 2007. Lexical Functions. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (eds.), *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung = Phraseology: An international handbook of contemporary research*, 118–131. Berlin & New York: Walter de Gruyter.
- Mel'čuk, Igor A. 2012. Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology* (3) 31–56. Berlin & New York: Mouton de Gruyter.
- Mollin, Sandra. 2009. Combining corpus linguistic and psychological data on word co-occurrences: Corpus collocates versus word associations. *Corpus Linguistics and Linguistic Theory* 5(2). 175–200.
- Nordquist, Dawn. 2004. Comparing elicited data and corpora. In Michel Achard & Suzanne Kemmer (eds.), *Language, culture and mind*, 211–223. Stanford, CA: CSLI Publications.
- Padučeva, Elena V. 2003. Taksonomičeskaja kategorija kak parametr leksičeskogo značenijskogo glagola [Taxonomic category as a parameter of the lexical semantics of the verb]. *Russkij jazyk v naučnom osveščanii* 2. 192–216.
- Padučeva, Elena V. 2004. *Dinamičeskie modeli v semantike leksiki* [Dynamic models in lexical semantics]. Moskva: Jazyki slavjanskoj kul'tury.
- Peškovič, Aleksandr M. 2001 [1914]. *Russkij sintaksis v naučnom osveščanii* [The syntax of Russian from a scientific perspective]. 8th edn. Moskva: Jazyki slavjanskoj kul'tury.
- Taylor, John R. 1998. Syntactic constructions as prototype categories. In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, 177–202. Mahwah, NJ & London: Lawrence Erlbaum.
- Tixonovič, Anžella N. 2009. Leksičeskaja funkcija Magn v sovremennom ruskom jazyke: korpusnoe i eksperimental'noe izučenie [Lexical Function Magn in present-day Russian: A corpus-based and experimental study]. Sankt-Peterburg: Rossijskij gosudarstvennyj pedagogičeskij universitet im. A. I. Gercena dissertation.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam & Philadelphia: John Benjamins.
- Uryson, Elena V. 2000. Ponjatie normy v metajazyke sovremennoj semantiki: parametry čelovečeskogo tela c točki zrenija russkogo jazyka [The notion of norm in the metalanguage of present-day semantics: Parameters of human body from the point of view of Russian]. In *Slovo v tekste i v slovare. Sbornik statej k 70-letiju akademika Ju. D. Apresjana*, 243–252. Moskva: Jazyki ruskoj kul'tury.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zaliznjak, Anna A. 1998. O meste vnutrennej formy slova v semantičeskom modelirovanii [Inner form of the word in the semantic modeling]. In *Dialogue'98: Computational linguistics and its applications*, 61–68. Kazan: Heter.
- Žolkovskij, Aleksandr K. & Igor' A. Mel'čuk. 1965. O vozmožnom metode i instrumentax semantičeskogo sinteza [Towards a feasible method and tools of semantic synthesis]. *Naučno-texničeskaja informacija* 5. 23–28.
- Žolkovskij, Aleksandr K. & Igor A. Mel'čuk. 1967. O semantičeskom sinteze [Towards semantic synthesis]. *Problemy kibernetiki* 19. 177–238.

Corpora

Russian National Corpus—RNC (<http://www.ruscorpora.ru>)

