

A.-J. Kyröläinen
(Turku)

Low-Frequency Constructions and Salience: A Case Study on Russian Verbs of Motion of Dative Impersonal Construction Type

Introduction

It is a basic tenet in cognitive and functional linguistics that language use shapes and forms “grammar”. However, the exact mechanisms which constitute this phenomenon are under constant debate; one of the most prominent variables is frequency of occurrence, and particularly high frequency items. The concept of frequency of occurrence is based on the underlying tenet of repetition; the activation of a linguistic item affects the conceptual representation. Thus, the repeated occurrences of the same linguistic item will ultimately lead to a stronger impact on the conceptual representation than isolated instances. Furthermore, the activation of linguistic items gives rise to general structures producing a conventional set of general and specific structures that allows the expression of both conventional and novel ideas (see, e.g. Bybee 2007, 1985; Bybee & Hopper 2001; Langacker 1991, 1999). The usage-based approaches to language have in common that they postulate grammatical structure to be the result of a routinization process which reduces cognitive effort in language production.

Generally speaking, frequency of occurrence is often taken to be relatively unproblematic in corpus linguistics. The dramatic increase in speed and access to large language corpora can be seen as a contributing factor. The utilization of frequency effects, based solely on absolute frequency as an explanatory factor for natural language, can be taken as presenting an oversimplified enterprise. The primary status of absolute frequency can be questioned by drawing an analogy between taboo concepts and absolute frequency. Corpus linguistics being primarily a methodological approach from form to meaning would posit that taboo concepts would not have any substantial significance if the primary explanatory factor were placed on absolute frequency. The analogy also extends to low-frequency linguistic items. In this approach to linguistic representation, the correlation between frequency and entrenchment

would be taken to form a one-to-one relationship. Intuitively, this claim would not seem appealing.

The concept of saliency is postulated to influence conceptual representation along the typical frequency effects. Furthermore, it is quite commonly argued that low frequency items can be salient (cf. Hoffmann 2004). The concept of saliency or prominence is often evoked in approaches driven by psychological adequacy (e.g. Langacker 1999; Dik 1981).

The relationship between frequency, to be more precise, low-frequency and saliency is explored in this paper¹. I will build my argument step by step: the concept of saliency is divided into two concepts: cognitive and linguistic saliency; the premise concerns the latter in this paper. Further, it is argued that at least four properties play a decisive role in determining linguistic saliency: productivity, idiosyncrasy, complexity of form, and complexity of meaning. The empirical body of data consists of one Russian impersonal construction type, the dative impersonal construction. In the first section, the concept of saliency is elaborated more precisely. The second section presents the Russian impersonal construction type under investigation. In the third section, the data are discussed in detail. The following sections are thereafter used to analyze the corpus-based data, and to discuss the findings in relation to the concept of linguistic saliency. In addition, the dative impersonal construction type has not been studied with a corpus linguistic approach before, at least to my knowledge. The data are fully based on corpus material, thus allowing the evaluation of claims made in previous studies on this particular construction type.

The concept of saliency

The concept of saliency is often evoked in cognitive and functional approaches to language. The saliency argument posits that certain concepts are taken to be more central in the mental representation relative to others. This notion of saliency is typically denoted with the term cognitive, i.e. some concepts are taken to be more central in the mental organization of language. However, although the concept of (cognitive) saliency is easy to make, it is typically left unspecified. Moreover, it is even harder to verify: how to objectively measure the saliency of certain concepts in mental representation. To complicate matters even more the concept of saliency is applied to a number of different phenomena: semantic roles, animacy hierarchies and, generally, language acquisition, among others (cf. Deumert 2003; Goldschneider & Dekeyser 2005; Levin & Rappaport Hovav 2005; Newmeyer 2002). Langacker (1988: 75) acknowledges the difficulties associated with the concept of saliency but he argues that the problems of verification do not merit the removal of the concept itself. He continues the argument that theoretical linguistic

cannot always operate on concepts based on empirical facts but some hypothetical concepts have to be introduced in order to seek explanatory adequacy (cf. Chomsky 2002; Popper 1968: 47, 51-52). The commonality of different approaches to salience is motivated by factors concerning language users; for instance, Newmeyer (2002: 50-51) outlines the general principle of salience in terms of centrality to human experience. Thus, the motivation for salience does not come from the language itself but is external to it. This distinction allows the separation of cognitive salience and linguistic salience from one another. The former pertains to the mental representation and the latter to the possible observable effects on language at the structural and semantic level.

The concept of salience offers the possibility to account for low-frequency items in a more natural manner. Low-frequency items may appear infrequently simple due to their inherent nature of expressing relatively rare concepts. However, the infrequent item may be the preferred choice of expression of the given concept and thus be more deeply entrenched than a more frequent item, which would be a nonpreferred choice of expression of the same concept (Hoffmann 2004: 190-191). As a result of this, the focus shifts away from pure absolute frequency of occurrence and offers a less mechanistic model of mental organization. In doing so, the role of mental representation is not taken to merely reflect repetition, but also the role of the language users becomes more active. The incorporation of salience does not contradict the findings of frequency based accounts (e.g. Bybee 2007, 1985; Goldberg 2006, 2003).

Intuitively, the concept of salience of a certain concept relative to others seems natural. However, the main questions still remain: how to account for linguistic salience, and which phenomena in a linguistic structure may possess diagnostic properties for salience. The Russian dative impersonal construction type may be considered as an ideal candidate for exploring the possible range of linguistic salience effects. The construction type satisfies the prerequisites of an ideal candidate: it is used to refer to human experience and it is primarily encoded with human referents. In addition, the construction type is infrequent. In this paper, four hypotheses are postulated and tested regarding the diagnostic properties of linguistic salience: if the construction type possesses salience, the following conditions should be met and thus should function as indicators of preferred choice of expression: productivity, idiosyncrasy, complexity of form, and complexity of meaning. Interestingly, the proposed indicators are also used in language acquisition studies as a broad conceptualization of salience (cf. Goldschneider & Dekeyser 2005). Before turning to the indicators, the Russian dative impersonal construction type is defined, and the data used in this paper are discussed in detail.

Russian dative impersonal construction type defined

In this corpus-based study, I explore the usage of the Russian dative impersonal construction type which is only one of the possible impersonal constructions in Russian. Before defining the construction type in detail, a look is taken at the previous studies on impersonal constructions in Russian. According to the standard definition, the impersonal constructions do not have a nominative subject; consequently, there is no syntactic agreement between subject and predicate which are presented by a non-nominative NP and a verb form either in third singular in the present tense or neuter singular in the past tense. (e.g. Arutjunova 1999; Bulygina & Šmelev 1997; Galkina-Fedoruk 1958; Leinonen 1985; Pariser 1982; Šahmatov 1925; Vinogradov 1972; Janko-Trinickaja 1962) The standard definition gives rise to numerous different types of impersonal constructions in Russian and only some of them are marked with the reflexive coding *-sja* (*-ся*).

The complexity of the impersonal category is also reflected in the classification of possible types in Russian. Šahmatov (1925) gives only one general category which is later divided into two classes by Vinogradov (1972). One of the most extensive studies on Russian impersonal constructions is presented by Galkina-Fedoruk (1958). She separates reflexive impersonal constructions, whether a verb has a reflexive suffix or not, from other possible impersonal types, and classifies and distinguishes the reflexive marked verbs into 10 groups based on lexical-semantic criteria. However, she does not offer clear-cut criteria for the classification and, furthermore, the relations between different lexical-semantic criteria are not elaborated; for instance Galkina-Fedoruk classifies *хотелось* ‘want (impersonal)’ and *спится* ‘sleep (impersonal)’ into the same group, although they have different markedness patterns.

Gerritsen (1990: 124 ff) argues that the previous studies on Russian impersonal constructions do not provide sufficient semantic features which would motivate the use of the attested patterns in Russian. He classifies impersonals into two broad classes: impersonal improper and impersonal proper. Moreover, the classification stems from the semantic properties of individual verbs which are in direct contrast to the framework adopted in this study, namely, the concept of construction in the sense of learned pairings of form and meaning. The concept of construction allows the separation of individual lexical items from general patterns and simultaneously maintains strict taxonomical criteria.

Although constructions as such have held a central position in linguistics and philosophy since Aristotle, the basic tenet in constructivist approaches to linguistic structure is the separation between lexical items and constructions. Constructions are defined as learned generalizations over surface structures.

The crucial assumption is that the constructions themselves are stored in the mental representation in the same way as lexical items. Thus, constructions are taken to exist in natural language independently of lexical items (cf. Croft & Cruse 2004; Goldberg 2006, 2003, 1995; Kay & Fillmore 1999). Goldberg (2006, 1995) posits that individual verbs involve roles which are to be associated with the frame semantics of the verb, i.e. encyclopaedic knowledge, for example the verb *мыть* ‘wash’ consists of such participant roles as *мойщик* ‘washer’ and *мытое* ‘washed’. The roles involved in constructions are called argument roles, and they are generalizations over participant roles of a lexical item. In this case, the argument roles are agent and patient. The participant roles of a verb and the argument roles of a construction are integrated in actual use, i.e. a correspondence is posited between participant and argument roles.

Recognizing the independent nature of constructions allows us to state generalizations directly from the surface structure. Moreover, the dative impersonal construction type naturally forms a specific pattern which is observable on any surface manifestation². Based on the surface generalizations, the construction has the following properties: dative experiencer, adverbial modifier, including the negative particle *не* ‘not’, and a verb marked with the reflexive marker³. The modifier is obligatory in the declarative clause but, in the case of the interrogative clause, the interrogative functions as a modifier, even in indirect questions. The following examples demonstrate the different possible patterns of this construction type:

1. *Но мне почему-то бега-ет-ся на*
 But I.DAT some reason run-NON.3.SNG.PRE-RM PR
таком покрыт-ии лег-че
 that pavement-PREP easily-COMP
 But for some reason it’s easier for me to run on that kind of pavement
 (Волжский Комсомолец (Самара), 04.02.2005).

2. [...] *как ему лета-ет-ся на наш-ем*
 how he.DAT fly-NON.3.SNG.PRE-RM PR our-PREP
«Союз-е»
 Soyuz-PREP
 [...] how he feels about flying with our Soyuz (spacecraft)
 (Областные вести (Волгоград), 03.02.2006).

Furthermore, the event is construed as state in this construction type. This can be tested with negation. In Example 3, the negation is used as a modifier and it depicts the event from the point of view of the experiencer (cf. Wierzbicka 1988: 230-231). The experiencer lacks only the state of running, but the state

in the following section. First, I will give a short introduction to the verbs of motion in Russian and then describe the used sampling frame and the reasons for choosing Integrum as the source of the base material. This is directly related to the question of corpus representativeness.

Verbs of motion					
	Intransitive			Transitive	
	Non-directional	Uni-directional		Non-directional	Uni-directional
'run'	бегать	бежать	'drive', 'run errands'	гонять	гнать
'wander'	бродить	брести	'roll'	катать	катить
'travel', 'ride'	ездить	ехать	'carry'	носить	нести
'climb'	лазить	лезть	'pull', 'drag'	таскать	тащить
'fly'	летать	лететь	'lead'	водить	вести
'swim', 'float'	плавать	плыть	'convey', 'transport'	возить	везти
'crawl'	ползать	ползти			
'go', 'walk'	ходить	идти			

Table 1: Russian unmarked verbs of motion

The pairwise occurring verbs of motion constitute a small and closed class in Russian. A full list of the unmarked verbs is presented in Table 1⁶. There are no generally accepted terms for the pairs: perhaps the more widely used terms in English are determinate and indeterminate. I follow the terminology used by Nessel (2000): unidirectional corresponds to determinate and non-directional to indeterminate. This terminology is also closer to the one used in the Academy Grammar (Švedova et al. 1982: 590-591). In addition, the adopted terminology reflects the semantic difference between the pairs. Unidirectional verbs profile an event which denotes motion in one direction and towards an endpoint. The non-directional verbs have wider semantics scope, i.e. they profile a range of different motion patterns. Typical instances are: firstly, general motion, round about, secondly, motion towards an endpoint and back again; thirdly, a general ability to perform the motion in question. Thus, the verbs in question form a semantic opposition where the unidirectional verbs of motion are specified in relation to directionality and the non-directional verbs are not (cf. Jakobson 1989: 47; Nessel 2000). In addition to the concept of directionality, the Russian verbs of motion can be further categorized on the basis of transitivity (cf. Hopper & Thompson 1980). This last feature directly relates to the formation of the dative impersonal construction. Typically, it is argued that

high transitive verbs, which always imply an object, cannot be used in this construction (Israeli 1997 cf. Arutjunova 1999: 803). Furthermore, the transitive verbs of motion also possess a form marked with the reflexive suffix. The relevance of both of these features is discussed in the following section.

The data for this study are compiled using the data base of Integrum (www.integrum.ru). Integrum is in a sense a monitor corpus, i.e. new material is added on an almost daily basis, and thus the size of the data base is constantly increasing. The data base mostly consists of Russian newspapers but other sources are also included: e.g. literature and internet forums. Integrum is comparable to The Global English Monitor Corpus (www.corpus.bham.ac.uk/ccl/global.htm). However, Integrum is not designed for linguistic purposes: there is no aim at representativeness or at a balance between the different sources of material⁷. As the data base of Integrum is not designed for linguistic purposes, the material is not annotated. At the moment, Integrum offers a fairly robust implementation of regular expressions which can be used to retrieve data. A precise search is used in this study to access the possible instances of this construction type. The used sampling frame consists of the verb of motion in third singular in the present tense and singular neuter in the past tense plus the reflexive marker. The sampling frame is also restricted on the basis of different registers in order to maintain stability between different sources (cf. Biber 1993). The following sources are included in the sampling frame: computer and foreign publications, Internet media, journals and magazines, media monitoring, news agencies, newspapers, regional newspapers, TV and Radio monitoring and world news agencies. Every retrieved hit is processed to determine its relevance for this construction type, the removal of duplicates and so on⁸. Using Integrum as the source of the base material, it was possible to extract a total of 2058 instances of use with the intransitive verbs of motion in this construction type. However, the retrieved instances cannot be normalized to a common base, i.e. it is not possible to make any claims based on absolute frequency of occurrence in the Integrum database.

Methodologically, there is a constant debate in the field of corpus linguistics concerning the status between monitor and balanced sample corpora designs (cf. Tognini-Bonelli 2001; Sinclair 2004 and references therein). The main argument concerns the sample sizes: the general size of a single sample as well as the overall number of samples in a corpus. They are claimed to be too small to give an accurate representation. The size of the modern balanced sample corpora is fairly extensive as is the case with The Russian National Corpus (www.ruscorpora.ru) consisting of ca. 140 million words. In addition, Biber (1993) has demonstrated that even 1000 word samples are stable enough for establishing typical patterns in language use. However, the situa-

tion is changing radically, as is also noted by Biber, in the case of low-frequency items.

Intransitive verbs of motions				
Non-directional			Unidirectional	
	Frequency			Frequency
бегаётся	2	'run'	бежится	2
бегалось	3		бежалось	4
бродится	1	'wander'	бредётся	0
бродилось	1		брелось	0
ездится	0	'travel, ride'	едётся	2
ездилось	2		ехалось	6
лазится	0	'climb'	лезётся	0
лазилось	0		лезлось	0
летается	1	'fly'	летится	0
леталось	1		летелось	1
плавается	3	'swim', 'float'	плавается	0
плавалось	1		плылось	1
ползается	0	'crawl'	ползётся	0
ползалось	0		ползлось	0
ходится	2	'go' 'walk'	идётся	6
ходилось	4		шлось	7

Table 2: The Russian National Corpus and the intransitive verbs of motion

This is evident if the Russian National Corpus is used as the base material; most of the verbs of motion are not attested in this construction type based on raw frequencies in Table 2⁹. Cumulative representativeness is needed in order to establish extremely infrequent items.

Productivity and frequency effects

Productivity can be taken as a central property of salience effect. If the construction in question is the preferred choice of expression, the construction should be applicable to a number of different verbs. Indeed, it is generally accepted that the Russian dative impersonal construction is productive but its limits are unknown, i.e. what is the exact semantic scope of the construction (see, e.g. Apresjan 2005: 8-9; Gerritsen 1990; Israeli 1997). While no evidence is provided to support the productivity of this construction type, it can be intuitively sustained by using ad hoc formations.

In usage-based models, the productivity of a certain pattern is taken to correlate with type frequency. Type frequency is measured in relation to the number of items which instantiate that particular pattern. A certain pattern which is applicable to a number of different instances is more likely to be applied to

new instances. In this manner, the productivity of a certain construction is measured in relation to different items subsumed under it. (e.g. Bybee 2007, 1985; Croft & Cruse 2004; Goldberg 2006, 1995). Measuring productivity solely in this manner skews the notion of productivity towards high frequency constructions. According to this definition, the primary motivation of productivity is frequency: a certain pattern is productive to the extent of being applicable to a number of different instances. For a corpus-based study, an extremely large sample would have to be extracted in order to define productivity in terms of type frequency if the main interest is to determine whether a low-frequency pattern is indeed productive or not. Furthermore, there appears to be no qualitative difference between high- and low-frequency types based on this notion of productivity: both types have more or less the same inner structure consisting of relatively few high-frequency tokens and a relatively large number of low-frequency items as can be seen in Table 3.

A more sensitive approach to low-frequency patterns is applied in this paper in order to determine the productivity. Osherson et. al (1990 cf. Goldberg 2006: 99-100) demonstrated that speakers more readily apply a pattern to new instances if the target item is related to the category known to exhibit the source pattern. In a linguistic application, this would allow the determination of productivity in terms of coverage within a certain semantically related category. This is why the verbs of motion are selected as the source of data. The verbs of motion are closed class verbs in Russian and obviously semantically related to each other. Now the productivity hypothesis can be elaborated further: if the construction type is productive it should possess a high level of coverage within the verbs of motion class, i.e. most if not all of the verbs of motion should manifest this construction type. Based on the Integrum database, it is possible to verify the claim of productivity in terms of coverage. The data are given in Table 3.

Intransitive Verbs of motion				
Non-directional			Unidirectional	
	Frequency			Frequency
бегаётся	83	'run'	бежится	245
бегалось	36		бежалось	472
бродится	7	'wander'	бредётся	2
бродилось	4		брелось	2
ездится	148	'travel', 'ride'	едётся	85
ездилось	172		ехалось	59
лазится	1	'climb'	лезётся	0
лазилось	0		лезлось	1
летается	108	'fly'	летится	68
леталось	105		летелось	26

Intransitive Verbs of motion				
Non-directional			Unidirectional	
	Frequency			Frequency
плавается	52	'swim', 'float'	плывется	0
плавалось	28		плылось	59
ползается	1	'crawl'	ползется	0
ползалось	1		ползлось	0
ходится	69	'go', 'walk'	идется	171
ходилось	18		шлось	35

Table 3: Integrum and intransitive verbs of motion and productivity in terms of coverage

The question of representativeness is generally apparent for low-frequency constructions, and it becomes even more crucial when a low-frequency construction, i.e. dative impersonal, is combined with low-frequency items as is the case with certain verbs of motion, namely, 'climb' and 'crawl', in Russian. This is evident when the figures are compared in Table 3. Some of the verbs are not attested in this construction type in Integrum but using Google and Yandex it is possible to find instances of use with these verbs. This confirms the full coverage of the construction type within the verbs of motion class.

It is claimed that telic verbs cannot appear in this construction (see, esp. Israeli 1997: 133-135). This would lead to the conclusion that the unidirectional verb of motions would not be attested in this construction type, at least with an overtly coded endpoint. However, when the corpus linguistic approach is taken seriously it is possible to evaluate the claim of the previous assertion. If The Russian National Corpus were used as the base material, it would certainly qualify to support this notion. However, like any negative evidence the absence of a certain form cannot be taken alone to support a claim (e.g. Penke & Rosenbach 2004). We simply cannot know whether the chosen base material is actually representative enough for an extremely infrequent construction to be included in the data. The failure to attest certain linguistic items / patterns may be due to some principled reason or coincidence, for example, the base sample may not consist of a large enough data set. Thus, the use of negative evidence as supporting evidence for a theoretical claim is not sufficient. Typically, the negative evidence is supported with additional evidence which is not directly related to the phenomenon in question. Indeed, this is the case with the unidirectional verbs of motion. The use of unidirectional verbs is prohibited in this construction if they possess a telic reading (e.g. Israeli 1997: 133-135; Pariser 1982: 72, 78-81 cf. Gerritsen 1990: 180-181). This hypothesis is wider in scope, including other than the verbs of motion, and thus it presents a situation which is applicable to a larger data set. In contrast, when Integrum is chosen for the base material the situation changes dramatically; not only are the unidirectional verbs of motion attested in this

construction type but they also possess a high token frequency. The unidirectional verbs of motion are used in this construction type with locative endpoints as in examples 5-8. Thus, it cannot be stated that the construction type would simply override the unidirectional, telic, meaning of the verbs.

5. [...] *от крепост-и до Советск-ой площад-и*
 PR stronghold-GEN PR Soviet-GEN square-GEN
беж-ит-ся быстр-ее.
 run-UNI.3.SNG.PRE-RM fast-COMP
 [...] it's faster to run from the stronghold to the Soviet Square
 (КузПресс (Новокузнецк), 04.07.2006).
6. [...] *во-вторых, домой всегда беж-ит-ся*
 secondly, home always run-UNI.3.SNG.PRE-RM
лег-че [...]
 easily-COMP
 [...] secondly, it's always easier to run home
 (Советская Сибирь (Новосибирск), 29.04.2003).
7. *После кружк-и пив-а домой Клав-е*
 PR mug-GEN beer-GEN home name-DAT
ш-л-о-сь весел-ее [...]
 walk-UNI. NEUT. PAST-RM cheerful-COMP
 For Klava it's more cheerful to run home after a pint [...]
 (Вятский наблюдатель (Киров), 01.12.2006).
8. [...] *до драмтеатра что-то не*
 PR theatre for some reason NEG
ид-ет-ся
 walk-UNI.3.SNG.PRE-RM
 [...] for some reason I don't feel like walking to the theatre.
 (Новая новгородская газета, 07.06.2006).

The verbs have a stative reading which is in accordance with the general semantics of this construction type, referring to the state of the experiencer. The unmarked verb form contains a starting and an endpoint and also a path between these two points. The configuration as a whole can be integrated into the dative impersonal construction type and there is no semantic conflict. The reading can either be habitual as is the case in Example 7 or unidirectional as in Example 8.

Table 3 consists only of the intransitive verbs of motion. The transitive verbs of motion do not appear in this construction type. This can be taken as verifying the transitivity claim. However, it is possible to find instances of

use with typical transitive verbs in this construction type, if they are properly contextualized. Example 9 illustrates that even high-transitive verb, *стрелять* ‘shoot’, can be coerced into this construction type. The proposed transitivity rule appears to be prematurely proposed.

9. [...] *мне* *легче* *беж-ит-ся* *и*
 I-DAT easily-COMP run-UNI.3.SNG.PRE-RM and
мет-че *стреля-ет-ся*
 accurately-COMP shoot-3.SNG.PRE-RM
 [...] it’s easier for me to run and shoot more accurately
 («Завтра снова в дорогу», Минский курьер, 15.07.2005).

The transitive verbs of motion have a reflexive marked form, but these forms differ in meaning compared with the unmarked form, for example *нести* ‘carry’ and *нестись* ‘rush’. The transitive verbs of motion were tested using the following sampling frames: [1.person-dative] + [negation] + [verb-present-singular.3.-RM / past-neuter-RM] and [1.person-dative] + [adverb *хорошо* / *легко*] + [verb-singular-present.3.-RM / past-neuter-RM] in order to find possible uses of this construction type. If transitive verbs can be attested in this construction type, it could at least be assumed that examples of this construction type would be found with the transitive verbs of motion. However, the used sampling frames did not produce any relevant hits.

I am inclined to propose that the transitivity of the verbs of motion per se is not the decisive factor in this case, but that the already existing reflexive form restricts the use of this construction type. Within the usage-based framework, several possible motivations for the restrictive use of these transitive verbs of motion can be postulated: First, the reflexive marked forms possess a relatively high token frequency, and therefore they retain a certain level of autonomy (c.f. Bybee 2007). Second, as the verbs in question have a lexicalized meaning different from the unmarked form, this would lead to competition between the two forms¹⁰. It is reasonable to argue that the lexicalized reflexive form is more entrenched in Russian, i.e. it is not used as a compound form but instead as a single unit (e.g. Langacker 1991). This relative autonomy of the meaning would thus function as a conservative factor, preventing these verbs from being rendered in this construction type. Interestingly, Järvikivi et al. (2006) have shown in regards to Finnish suffix allomorphy for processing complex words that the frequency alone is not the decisive factor in language processing; the structural invariance also plays an important role. We can postulate the following hypothesis: the form of a given verb and its connection to productivity in this construction type are related to structural invariance. It follows from this that those verbs, which are closer to the invariance of the structure, are unlikely to appear in this construction type.

However, it is not possible to test this hypothesis further in this paper: a larger set of data would need to be gathered in to determine the scope of productivity in terms of structural invariance.

Collocates and the level of idiosyncrasy

According to the previous studies (see esp. Zolotova et al. 1998), the impersonal construction type is mostly used in conjunction with the negative particle *не*. This claim would allow further postulating that there is a high level of idiosyncrasy involved, or to be more precise, the construction type involves highly idiomatic patterns. This structural behaviour of this construction is tested with collocates, the recurrent co-occurrence of lexical items. Collocations of a certain linguistic item are readily identifiable on the vertical axis in concordance. (e.g. Tognini-Bonelli 2001). Collocations allow accessing the co-occurrence of the adverbs of manner in order to test the claim of the extensive use of the negative particle in this construction type.

The following criteria are used to establish the collocations. The minimal frequency of occurrence of a single lexical item is set at 10. In addition, the maximal length of collocations is used allowing five positions both to left and right, if possible, before a sentence break. The sampling frame consists of both present and the past tense forms of a single verb of motion. The collocation information is presented by combining the different sampling frames, i.e. tense and non-directional and unidirectional forms. Those verbs of motion which are infrequent i.e. have less than 10 hits in the data, are not listed in the collocation table; they do not meet the minimum requirement of the frequency of occurrence. The results of the study of collocations are given in Table 4.

	Number of verbs		Collocations						verbs
			легко 'easily'	легче 'more easily'	не 'not'	хорошо 'well'	лучше 'better'	тяжело 'difficult'	
бегаться / бежаться	'run'	836	195	82	218	85	15	53	15
ездиться / ехать	'travel', 'ride'	464	24		41	22			240
летаться / лететь	'fly'	307			134	18			29
плаваться / плыть	'swim', float	139	16		21				14
ходиться / идти	'go', 'walk'	293	23	15	207				80
Total		2039	258	97	621	125	15	53	378

Table 4: Integrum, collocations and the intransitive impersonal verbs of motion

The data themselves do not allow direct comparison of distribution patterns of collocates with the individual verbs of motion due to the difference in sample sizes, but Table 4 suggests that, although the frequency of occurrence of the negative particle is evident, there is no high level of co-occurrence of the negative particle in this construction type with the exception of the verb ‘go’, ‘walk’. Additionally, the verb ‘run’ appears with the adverb *тяжело* ‘difficult’ being the only verb possessing this collocate. Moreover, this is the only verb to show the full collocation set presented in Table 4. This fact is probably due to the representativeness of the sample size relative to the verbs of motion in the data. Furthermore, the collocation data would postulate that not only are certain idiomatic patterns, which are shared by most of the verbs of motion, characteristic of this construction type, but certain verbs may also possess individual, idiomatic behaviour. This would confirm the separation of lexical items from constructions. The construction type itself has idiomatic lexical patterns but verbs may also introduce their own set when used in the given construction¹¹.

Another idiomatic feature emerged when the collocations were analyzed with the adverbs of manner in this data, namely, the clustering of different verbs. At least with the verbs of motion, there is another idiomatic pattern involved when this construction type is used; not only are the adverbs of manner highly idiomatic but there also appears to be a clustering of verbs in this construction type. The verbs of motion form a chain consisting of other verbs which are used in the same construction type. Interestingly, the typical co-occurring verbs belong to the same frame as the pairing verbs of motion, for instance, *дышаться* ‘breathe’, *прыгаться* ‘leap’ and *шагаться* ‘step’. This is in accordance with the frame-semantic approach used in this paper (cf. Fillmore 1982). The verbs of motion profile a frame which may be referred to as bodily activity. The construction type is used to construe a certain event as a state of the experiencer in conjunction with the bodily activity frame. The activation of the bodily action frame enables the accessing of other verbs depicting motion, for instance, in Example 10. While the verbs denoting the bodily activity frame are the only one satisfying the frequency of occurrence criteria, the chain may also include verbs denoting some other frame as is the case in Example 11.

10. *Причин-а* *рост-а* *пострадави-их* *очевидна* —
Reason-NOM increase-GEN injured-GEN apparent —
гололед. *Не* *ид-ет-ся* *и* *не*
black frost-NOM NEG walk-UNI.3.SNG.PRE-RM and NEG
ед-ет-ся, *но* *зато* *прекрасно*
drive-UNI.3.SNG.PRE-RM but instead nicely
пада-ет-ся.
fall-3.SNG.PRE-RM

The reason for the increase in the injured is apparent — black ice. You don't feel like walking and driving, but instead slipping is easy (Саратов — Столица Поволжья (Саратов), 03.03.2005).

11. [...] *что здесь мне ле-гче дыш-ит-ся,*
 that here I.DAT easily-COMP breathe-3.SNG.PRE-RM
лег-че пиш-ет-ся, лег-че
 easily-COMP write-3.SNG.PRE-RM, easily-COMP
ход-ит-ся по земл-е.
 walk-NON.3.SNG.PRE-RM PR earth-PRE
 [...] that for me it is easier to breathe and write here and easier to walk
 on the earth
 (Полоса, 08.09.2004).

Complexity of form and complexity of meaning

The concept of complexity is approached in this paper in relation to both form and meaning. Although the complexity of form is connected with the complexity of meaning (cf. Haiman 1983; Nessel 2000), positing an iconic relationship between them, it is relative easy to find counter-examples to this; for instance, the complexity of form does not necessarily equate with the complexity of meaning or vice versa as is the case with the diachronic development of the Russian reciprocal pronoun, from a simple to a compound form. Thus, the iconic relationship is taken as the point of departure of the discussion in this section.

The concept of complexity of the form is defined in this paper through the markedness pattern of the construction type compared to the unmarked constructions. The complexity of the markedness pattern is due to several reasons. Firstly, the experiencer is encoded with an oblique case, the dative. This figures prominently in the structural complexity of the construction compared with the unmarked ones. Secondly, the adverb of manner is an obligatory argument in the construction type. This is not the case with the unmarked form where the adverb of manner is a facultative argument. Finally, the verb forms are marked with the reflexive marker contributing to the overall complexity of the form in this construction type.

The complexity of meaning and its connection with the complexity of the form comes from the adverb of manner. Although the obligatoriness of the adverb of manner contributes to the complexity of meaning, it primarily results from the different interpretation of the adverb of manner in the impersonal construction. According to Adamec (1973: 121-122; cf. Gerritsen 1990: 176-177) the adverbs of manner typically receive a subjective reading when used in this impersonal construction type. This is exemplified by Example 12, where the

adverb of manner does not depict the qualitative nature of the running; instead it is used for purposes of subjective evaluation. Moreover, the performance is depicted from the point of view of the experiencer in Example 12.

12. *На всех этапах Кубка на дистанции 1500 метров он был первым или*

In all stages of the Cup in the 1500 metres race he was either first or
вторым.
second.

Юноша признается, что 500 метров у него любимая дистанция,
The Youngster confesses that his favourite distance is the 500 metres,
10000 — легч-е всего беж-ит-ся.

10000-NOM — easily-COMP all in all run-UNI.3.SNG.PRE-RM

In all stages of the Cup he was either first or second in the 1500 metres race. The youngster confesses that his favourite distance is the 500 metres, although the 10 000 metres race is easiest
(Сельская жизнь, 04.12.2003).

Further complexity of meaning is due to the use of the negative particle. The negative particle has, to a certain extent, the same characteristics as the use of the comparative in this construction type. The comparative explicitly functions as a comparison between two states as is demonstrated in Example 12. The use of the negative particle introduces a subtle comparison between different states, i.e. the actual state of comparison does not need to be overtly expressed. Example 13 illustrates this. The subtle comparative interpretation rises especially when an adverb of either space or time is also used in the construction.

13. [...] *чтобы мы не чувствовали себя там чужеземцами.*

So that we would not feel like strangers there.

Но опять не ед-ет-ся

But again NEG travel-UNI.3.SNG.PRE-RM

[...] so that we would not feel like strangers there. But again we don't feel like travelling

(Тюменские известия, 20.07.2002).

In addition to this comparative reading, the negative particle may also introduce a normative reading, for instance the example 14 (Trub 1994: 48). The experiencer is unable to perform, although it is expected. The normative reading arises from this tension between the two states.

14. *Никак не пойму, чего им в будн-и не
езд-ит-ся?*
At all I don't understand why they.DAT PR week-PL.ACC NEG
езд-ит-ся?
ride-NON.3.SNG.PRE-RM
I don't understand at all why they don't feel like riding during the
week
(Вечерняя Казань, 17.06.2006).

Discussion

In this paper, I have tried to answer the question of what motivates the use of the Russian dative impersonal construction, although its frequency of occurrence is relatively low, in terms of linguistic saliency. The concept of linguistic saliency is outlined in terms of semantic and structural indicators, i.e. productivity, idiosyncrasy, complexity of form and complexity of meaning. These indicators may be taken to function as diagnostic tools for establishing the effects of saliency at the semantic and structural level. The proposed diagnostic tools are by definition broad but they are neither language- nor structure-specific. Therefore, they are applicable to a number of different languages. On the other hand, the relevance of these diagnostic tools needs to be further evaluated through comparison to other linguistic constructions, whether by means of corpus-based studies or experiments on test subjects. A good candidate for testing these diagnostic tools would be, for instance, a study on idiomatic expressions.

In addition, the proposed diagnostic properties allowed evaluation of the claims made in previous studies on the Russian dative impersonal construction type and they also provided the theoretical framework which served to establish the semantic and structural specifications of this particular construction type. The productivity of the construction type was verified by investigating it in terms of coverage. The verbs of motion were used to establish the coverage of the construction type within this closed class. Furthermore, this paper presented evidence that even the unidirectional verbs of motion with overly coded locative endpoints are actively used in contemporary Russian. In addition, the corpus-based study showed that the construction type has a high level of idiosyncrasy patterns in terms of lexical items, for instance, the adverbs of manner are fairly stable in actual language use.

The concept of complexity of form and meaning allowed analysis of both the structural markedness and the semantic peculiarities of the construction in question. The adverbs of manner typically receive a subjective reading when used in the construction. They depict the event in qualitative terms from the point of view of the experiencer. Furthermore, the use of the negative particle adds semantic complexity to the construction either by making a subtle com-

parison between two states, although the compared state is rarely overtly expressed in actual usage, or by adding a normative reading to the whole construction. In the normative, frame it is expected that the experiencer should perform but for reasons unknown he / she is incapable of doing so.

In addition, this paper presented methodological considerations on low-frequency constructions. The concept of corpus representativeness was touched on in conjunction with the use of negative evidence. Finally, this paper continued the long tradition of studying the impersonal construction types in Russian from a corpus linguistic point of view.

Abbreviations in glossing

ACC = accusative	GEN = genitive	PAST = past tense	PREP = prepositional case
COMP = comparative	NEG = negation	PLURAL = plural	RM = reflexive marker
DAT = dative case	NEUT = neuter	PR = preposition	SNG = singular
NON = non-directional verb of motion	NOM = nominative	PRE = present tense	UNI = unidirectional verb of motion

Notes

¹ I would like to thank Riitta Pyykkö, Krista Ojutkangas and Seppo Kittilä for their helpful comments on the earlier versions of this paper. Additionally, I would like to thank the anonymous reviewers for their insightful comments, which were used in preparing the final version of this paper. I, of course, assume responsibility for any errors that remain.

² Although the primary concern of this paper is not to offer a full-blown model of all the possible types of the Russian impersonal constructions, the concept of constructions, i.e. generalizations over surface patterns, allows the distinguishing of several different impersonal types which, in previous studies, are arbitrarily lumped either into one or different categories. The mode of description in argument construction grammar can be seen as an elaboration of the Russian structuralist tradition, especially Zolotova (1973). Furthermore, Leinonen (1985) classifies different types on the basis of sentence patterns, for example, experiencer, natural phenomena, state, infinitival, and existential sentences. I would take this to be compatible with the model proposed in this paper.

³ Gerritsen (1990: 191-192 and references therein) argues that the construction type can appear without the adverb of manner. At least, the data in this study do not support this claim.

⁴ The data in the on-going study are based on different corpus materials.

⁵ Ultimately, the question is directly linked to the theoretical mode of description: whether redundancy is allowed in the description or not. This question separates different cognitive approaches from each other. The matter is beyond the scope of this paper but a relevant discussion can be found in Croft — Cruse 2004; Goldberg 1995, 2006; Langacker 1991; Kay & Fillmore 1999.

⁶ The list of verbs of motion is in accordance with the Academic Grammar. However, there is a certain level of variation among scholars; others exclude certain verbs while others include some more. A detailed discussion of this matter is offered by Nessel (2000). The inclusion or exclusion of certain verbs has no bearing on the current topic.

⁷ MacEnery et al. (2006) offer a more elaborated discussion on the problematic nature of monitor corpora.

⁸ There were a fair number of duplicates. Some data corruption is expected with large data sets but it seems that the primary reason for this is the way of storing samples in Integrum. The majority of the duplicates were found in samples from forum discussions. It appears that if a new entry occurred in the same discussion thread, it would be stored separately on the next data retrieval session simultaneously leading to the copying of previous entries into the new sample.

⁹ I would like to thank the anonymous reviewers for referring to the variation of the verb 'climb'. In my future studies, the variation will be taken into consideration.

¹⁰ A similar claim can be found in Galkina-Fedoruk (1958: 168-169 cf. Arutjunova (1999: 802) and references therein). She states that verbs which have a reflexive marked form cannot appear in this construction type. At least to my knowledge, this claim has not been made to other studies.

¹¹ The adverbs of manner given in Table 3 correspond to those presented in previous studies (c.f., e.g. Gerritsen 1990: 186-187). This may suggest an even deeper level of entrenchment of these lexical items in this construction type.

Bibliography

- Adamec, P.: 1973, *Očerk funkcional'no-transformacionnogo sintaksisa sovremennogo russkogo jazyka*, Praha.
- Apresjan, Ju.D.: 2005, 'O moskovskoj semantičeskoj škole', *Voprosy jazykoznanija*, 1, 3-30.
- Arutjunova, N.D.: 1999, *Jazyk i mir čeloveka*, Moskva.
- Biber, D.: 1993, 'Representativeness in corpus design', *Literary and Linguistic Computing*, 8, 4, 219-241.
- Bulygina, T.V. & Šmelev, A.D.: 1997, *Jazykovaja konceptualizacija mira (na materiale russkogo jazyka)*, Moskva.
- Bybee, J.L. (ed.): 2007, *Frequency of Use and the Organization of Language*, Oxford
- Bybee, J.L.: 1985, *Morphology: A Study of the Relation between Meaning and Form*, Amsterdam.
- Bybee, J.L., Hopper, P.: 2001, 'Introduction to frequency and the emergence of linguistic structure', J. Bybee, P. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam / Philadelphia, 1-24.
- Chomsky, N.: 2002, *On Nature and Language*, Cambridge.
- Croft, W., Cruse, D.: 2004, *Cognitive Linguistics*, Cambridge.
- Deumert, A.: 2003, 'Markedness and salience in language contact and second-language acquisition: evidence from a non-canonical contact language', *Language Sciences*, 25, 6, 561-613.
- Dik, S.C.: 1981, *Functional Grammar*, Dordrecht.
- Fillmore, C.J.: 1982, 'Frame Semantics', The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, Seoul, 111-137.
- Galkina-Fedoruk, E.M.: 1958, *Bezličnye predloženiya v sovremennom russkom jazyke*, Moskva.
- Gerritsen, N.: 1990, *Russian Reflexive Verbs. In Search of Unity in Diversity*, Amsterdam.

- Goldberg, A.E.: 2006, *Constructions at Work. The Nature of Generalization in Language*, Oxford.
- Goldberg, A.E.: 2003, 'Constructions: a new theoretical approach to language', *Trends in Cognitive Sciences*, 7, 5, 219-224.
- Goldberg, A.E.: 1995, *Constructions. A Construction Grammar Approach to Argument Structure*, Chicago.
- Goldschneider, J.M., Dekeyser, R.M.: 2005, 'Explaining the "Natural Order" of L2 Morpheme in Acquisition in English: A Meta-analysis of Multiple Determinants', *Language Learning*, 55, 1, 27-77.
- Haiman, J.: 1983, 'Iconic and economic motivation', *Language*, 59, 781-819.
- Hoffmann, S.: 2004, 'Are low-frequency complex prepositions grammaticalized?', H. Lindquist, C. Mair (eds.), *Corpus Approaches to Grammaticalization in English*, 171-210.
- Hopper, P., Thompson, S.: 1980, 'Transitivity in grammar and discourse', *Language*, 56, 2, 251-299.
- Israeli, A.: 1997, *Semantics and Pragmatics of the "Reflexive" Verbs in Russian*, München.
- Jakobson, R.: 1989, *Russian and Slavic Grammar, Studies 1931-1981*, L.R. Waugh, M. Halle (eds.), Berlin / New York.
- Janko-Trinickaja, N.A.: 1962, *Vozvratnye glagoly v sovremennom russkom jazyke*, Moskva
- Järvikivi, J., Bertram, R., Niemi, J.: 2006, 'Affixal salience and the processing of derivational morphology: The role of suffix allomorphy', *Language and Cognitive Processes*, 21, 4, 394-431.
- Kay, P., Fillmore, C.J.: 1999, 'Grammatical Construction and Linguistic Generalization: The What's X Doing Y? Constructions', *Language*, 75, 1, 1-33.
- Langacker, R.W.: 1999, *Grammar and Conceptualization*, Berlin / New York.
- Langacker, R.W.: 1991, *Foundations of Cognitive Grammar. V.2: Descriptive Application*, Stanford.
- Langacker, R.W.: 1988, 'A view of linguistic semantics', Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, Amsterdam / Philadelphia, 49-90
- Leinonen, M.: 1985, *Impersonal Sentences in Finnish and Russian: Syntactic and Semantic Properties*, Helsinki.
- Levin, B., Rappaport Hovav, M.: 2005, *Argument Realization*, Cambridge.
- McEnery, T., Xiao, R., Tono, Y.: 2006, *Corpus-Based Language Studies*, London / New York.
- Nesset, T.: 2000, Iconicity and Prototypes: 'A new perspective on Russian verbs of motion', *Scando-Slavica*, 46, 149-154.
- Newmeyer, F.J.: 2002, 'Optimality and functionality: a critique of functionally-based optimality-theoretic syntax', *Natural Language & Linguistic Theory*, 20, 43-80.
- Osherson, D.N., Wilkie, O., Smith, E.E., Lopez, A., Shafir, E.: 1990, 'Category-based induction', *Psychological Review*, 97, 2, 185-200.
- Pariser, J.R.: 1982, *Dative-Reflexive Constructions in Contemporary Russian*, Dissertation, Los Angeles.

- Penke, M., Rosenbach, A.: 2004, 'What counts as evidence in linguistics? An introduction', *Studies in Language*, 28, 3, 480-526
- Popper, K.R.: 1965, *Conjectures and Refutations: The Growth of Scientific Knowledge*, New York / Evanston.
- Sinclair, J.: 2004, *Trust the Text: Language, Corpus and Discourse*, London.
- Šahmatov, A.A.: 1925, *Sintaksis ruskogo jazyka*, Vypusk 1, Leningrad.
- Švedova, N.Ju. et al. (eds.): 1982, *Russkaja grammatika*, T. 1, Moskva.
- Tognini-Bonelli, E.: 2001, *Corpus Linguistics at Work*, Amsterdam.
- Trub, V.M.: 1994, 'O kommunikativnyh aspektah otricanija kak negativnoj ocenki istinosti', *Voprosy jazykoznanija*, 1, 44-61.
- Vinogradov, V.V.: 1972, *Ruskij jazyk (grammatičeskoe učenie o slove)*, Moskva.
- Wasow, T. & Arnold, J.: 2005, 'Intuitions in linguistic argumentation', *Lingua*, 115, 11, 1481-1496.
- Wierzbicka, A.: 1988, *The Semantics of Grammar*, Amsterdam / Philadelphia.
- Zolotova, G.A.: 2000, 'Ponjatije ličnosti / bezličnosti i ego intepretacii', *Russian Linguistics*, 24, 2, 103-115.
- Zolotova, G.A.: 1973, *Očerki funkcional'nogo sintaksisa ruskogo jazyka*, Moskva.
- Zolotova, G.A., Onipenko, N.K., Sidorova, M.Ju.: 1998, *Kommunikativnaja grammatika ruskogo jazyka*, Moskva.