Threatening in Russian with or without -*sja*: *grozit'* vs. *grozit'sja*

Anonymous Authors

This article explores the two verbs *grozit'* and *grozit'sja*, that can both be translated as 'threaten'. We adopt a "local" approach and offer a thorough analysis of corpus data, which indicates that the two verbs are semantically and syntactically distinct, although they share a number of properties. We show that the two verbs collocate with different parts of speech and tend to occur in different syntactic constructions. *Grozit'sja* is typically used about interactions between two persons, while *grozit'* has a wider range of uses. This tendency has become more expressed over time. As for the meaning of the verbs, *grozit'sja* tends to express verbal threats, while *grozit'* often conveys non-verbal threats. On a more theoretical level, our study contributes to our understanding of the morpheme -*sja*. While labels like "reflexive", "middle" and "passive" are helpful as far as they go, we demonstrate how detailed studies of individual verb pairs (a "local" approach) may shed light on the complex syntactic and semantic properties of -*sja.* On the methodological level, our study underscores the value of corpus data for the study of -*sja*, both data from large internet corpora such as the Araneum Russicum Russicum Maius and the Russian National Corpus (RNC). While the former corpus enables us to identify general tendencies through collocations and semantic vectors, a smaller curated corpus like the RNC is suitable for detailed analysis of semantic and syntactic properties.

Keywords:
Reflexive, middle, passive, -*sja*, corpus, semantic vector, collocation, threat

1. Introduction: the problem

Notorious for its polyfunctionality, the morpheme -*sja* represents a classic descriptive and theoretical problem in Russian linguistics. What is the meaning of -*sja*? What is the semantic and syntactic effect of adding -*sja* to a verb? As is well known, -*sja* is attested as a marker of middle (or reflexive) in (1) and passive in (2), as well as in a number of related functions, for which a wide variety of classifications and terminologies exist.[1]

(1)   Francuzskij korol' Ljudovik XI **mylsja** pjat' raz v god. (*Znanie-Sila* 2003)
      'The French king Louis XI washed five times a year.'
(2)   Kak ob''jasnit' istoriju pojavlenija na territorii Rossii nemetskix kirx, kotorye **stroilis'**
      tevtonami […]? (Demskaja 2017)
      'How can we explain the emergence of German churches on Russian territory, that
      were built by the Teutons […]?'

   Traditionally, -*sja* has been described in terms of a list or, especially in cognitive linguistics, a network of related meanings or functions (see, e.g., Geniušienė 1987, Enger and Nesset

---

[1] All numbered examples are from the Russian National Corpus, available at www.ruscorpora.ru. For each example, we provide the year of publication, as well as the name of the author (for works of fiction) or periodical (for works of non-fiction). For the convenience of the reader, the relevant verb or construction is boldfaced. A note on terminology is necessary. Many researchers refer to examples like (1) as "reflexive", but following Kemmer (1993) and Enger and Nesset (1998) we prefer the term "middle voice" for examples with -*sja*, thus reserving the term "reflexive" for sentences with the pronoun *sebja* (e.g., *nenavidet' sebja* 'hate oneself').

1998, Goto and Say 2009 and Kyröläinen 2013). In order to shed light on this list or network, two approaches are conceivable. A "global" approach involves investigating and classifying a wide variety of verbs, while what we may call a "local" approach offers in-depth analyses of individual verbs. The two approaches are complementary, and in actual practice most researchers who offer "global" analyses of the system of -*sja* as a whole also to some extent provide "local" descriptions of individual verbs (e.g., Israeli 1997, Knjazev 2007 and Goto and Say 2009).

In the present study, we adopt a "local" approach where we focus on the two near synonyms *grozit'* and *grozit'sja*, both of which can be translated as 'threaten' (Glovinskaja 2004a-b). The two verbs can be attested in very similar syntactic environments. In (3) and (4), for instance, both verbs combine with a nominative subject representing the "threatener" (the person who carries out the threat), a noun phrase in the dative representing the "threatenee" (the person who is threatened), and an infinitive complement representing the action the subject threatens to carry out:

(3)   Skol'ko raz Carevskij i Vevers **grozili mne sostavit'** protokol o moix popytkax "diskreditirovat' rukovodstvo obkoma [...]". (Ginzburg 1967)
       'How many times didn't Carevskij and Vevers threatened me to report my attempts to "discredit the leadership of the regional committee".'
(4)   [V]y časten'ko **grozilis' Čebakovu ujti** k svoemu professoru [...]. (A. & G. Vajner 1987)
       '[Y]ou often threatened Čebakov to go to your professor […].'

In examples like (3) and (4), *grozit'* and *grozit'sja* may be used interchangeably without clear semantic differences. We must therefore ask: What is the meaning of -*sja* in *grozit'sja*? What is the effect of adding -*sja* to *grozit'*? While several researchers have provided insightful analyses (e.g., Gerritsen 1990 and Israeli 1997), we are not aware of extensive investigations of data from large electronic corpora, using the methodologies of contemporary corpus linguistics. The present study aims at filling this knowledge gap.

Besides offering an analysis of *grozit'* and *grozit'sja* that has implications for our understanding of -*sja* in general, we address the culturally and linguistically important concept of "threat" that has received considerable attention in general linguistics in recent years. Cognitive linguists have discussed the verbs for threatening in English, Dutch, and Spanish (cf., e.g., Langacker 2000, Verhagen 1995 and Cornillie 2004). Examples like *The incident threatened to ruin his chances* (Verhagen 1995: 111) are argued to involve a high degree of "subjectification", whereby the likelihood of the relevant event (e.g., *to ruin his chances*) receives a positive or negative evaluation by the speaker (Cornillie 2004).

Experts on grammaticalization have been interested in verbs for threatening, since they represent a grammaticalization path from examples like (3) and (4), where a person promises to harm another person, to more abstract examples such as *The Australian dollar threatens to fall below 72 cents* (Narrog and Heine 2021: 32, see also Heine and Miyashita 2007 and 2008). Although subjectification and grammaticalization are not central topics of the present study, we note that Russian is of particular interest for linguistic investigations of threats since Russian has more than one morphologically related verb for 'threaten'. In addition to *grozit'* and *grozit'sja,* Russian also has the verb *ugrožat'* 'threaten', as well as a number of perfective verbs with prefixes, such as *prigrozit'* 'threaten'. In the present study, we limit ourselves to *grozit'* and *grozit'sja*, which are relevant for the study of -*sja*.

The contribution of our study can be summarized as follows. First, we show that *grozit'* and *grozit'sja*, although they show some degree of overlap, are syntactically and semantically

distinct. Second, our study illustrates the value of a "local approach" to -*sja.* Although simple labels like "middle voice" and "passive" are useful as far as they go, we also need detailed analyses of individual verbs in order to pinpoint all the idiosyncratic and unpredictable properties of -*sja.* Third, on the methodological level, our analysis indicates the usefulness of investigating semantic vectors and collocations in large internet corpora. However, at the same time we show that detailed analysis of individual examples from curated and balanced corpora is also required. Last but not least, our analysis demonstrates that Russian, like other European languages, has abstract examples that deviate from the prototypical situation where one person promises to do harm to another person. Interestingly, this mainly applies to *grozit'*, while *grozit'sja* is more likely to be used about prototypical threats.

Our argument is structured as follows. Sections 2 and 3 are devoted to semantic vectors and collocations in a large internet corpus (Araneum Russicum Russicum Maius).[2] In sections 4 and 5, we turn to data from the Russian National Corpus and consider argument structure constructions. Sections 6 and 7 concern the meaning of the two verbs under scrutiny, before we turn to the meaning and functions of -*sja* in section 8. Section 9 summarizes our findings.

2. Semantic vectors – how similar are *grozit'* and *grozit'sja*?

As a first step in our attempt at teasing apart the meanings and functions of *grozit'* and *grozit'sja*, we use semantic vectors (word embeddings), a method that has been gaining importance in corpus studies in recent years. As we will see, *grozit'* and *grozit'sja* do not come out as close relatives, but both are indirectly related through their common relative *ugrožat'*, which also means 'threaten'.

The idea behind semantic vectors is the Distributional Hypothesis that words with similar meanings tend to occur in similar contexts. If you happen not to know the word *sriracha* but notice that it shows up in texts together with *hamburger* in much the same way as, say, *ketchup*, *aioli*, *mayonnaise* and *béarnaise*, you might correctly guess that *sriracha* is a sauce that goes well with hamburgers. While the Distributional Hypothesis goes back at least to the 1950s (Joos 1950, Harris 1954, Firth 1957), it was only with the advent of large electronic corpora that it was possible to make real use of it. Combined with large corpora, semantic vectors offer enormous power to Natural Language Processing, as pointed out by Jurafsky and Martin (2021). It is possible to calculate a vector for each word based on all the contexts where it is attested in a corpus. The vector of each word can be represented as a point in a multidimensional space, where similar words are located close to each other.

The Araneum Russicum Russicum Maius corpus, a large internet corpus containing 1.2 billion Russian word tokens includes a function that enables us to investigate the similarity of words by means of semantic vectors.[3] For each word one searches for, the corpus returns a list of the 25 most closely related words, where "related" means that they occur in similar contexts in the corpus. Table 1 provides the lists for *grozit'*, *grozit'sja*, as well as the third imperfective verb for 'threaten', *ugrožat'*. As shown, the list for *grozit'* does not contain *grozit'sja*, and the list for *grozit'sja* does not contain *grozit'*. However, both lists include *ugrožat'*, and the list of *ugrožat'* contains both *grozit'* and *grozit'sja*. In other words, the

---

semantic vectors from the Araneum corpus indicate that *grozit'* and *grozit'sja* are related, but only indirectly through *ugrožat'*. The annotated gnuplot in Figure 1 visualizes this. This plot distributes the relatives of *ugrožat'* in a two-dimensional space. As shown, the upper left part of the space is inhabited by nominal forms (nouns and participles), while the lower left portion of the diagram involves verbs. *Grozit'* and *grozit'sja* are both related to *ugrožat'*.

Grozit'	Grozit'sja	Ugrožat'
0.000, грозить	0.000, грозиться	0.000, угрожать
0.488, **угрожать**	0.295, грозилась	0.470, угрожающего
0.536, чреватый	0.377, пригрозили	0.488, **грозить**
0.562, грозящее	0.384, пригрозить	0.519, угроза
0.601, грозящий	0.406, пригрозила	0.520, угрожающий
0.632, грозящего	0.510, грозясь	0.559, угрожавшего
0.635, грозящих	0.577, пообещать	0.609, **грозящих**
0.640, грозящий	0.611, обматерил	0.620, пригрозить
0.644, повлечь	0.612, выгнать	0.620, шантажировать
0.662, чреваты	0.613, отместка	0.626, шантажировали
0.666, грозивший	0.613, вознамериться	0.645, **грозиться**
0.668, обернуться	0.615, припугнул	0.646, угрожаем
0.683, грозившая	0.628, засудят	0.647, **грозящего**
0.687, караться	0.630, засудит	0.655, запугивал
0.691, неминуемый	0.636, посметь	0.663, запугивать
0.693, штраф	0.645, **угрожать**	0.671, угрожай
0.693, пригрозить	0.649, шантажировали	0.675, **грозящее**
0.696, спровоцирует	0.650, пожалуется	0.677, обеспокоить
0.702, влечь	0.651, наорал	0.684, **грозящий**
0.712, грозящем	0.662, вышвырнуть	0.684, **грозящий**
0.714, нанесло	0.665, нажаловалась	0.685, опасаться
0.716, обеспокоить	0.666, запугивал	0.689, угрожающем
0.716, пригрозили	0.666, шантажировать	0.693, угрожающего
0.717, наказываться	0.666, подговорили	0.694, угрожавшую
0.717, грозившей	0.667, накажут	0.694, пригрозили

Table 1: The 25 most closely related words to grozit', grozit'sja, and ugrožat' based on semantic vectors from the Araneum Russicum Russicum Maius corpus
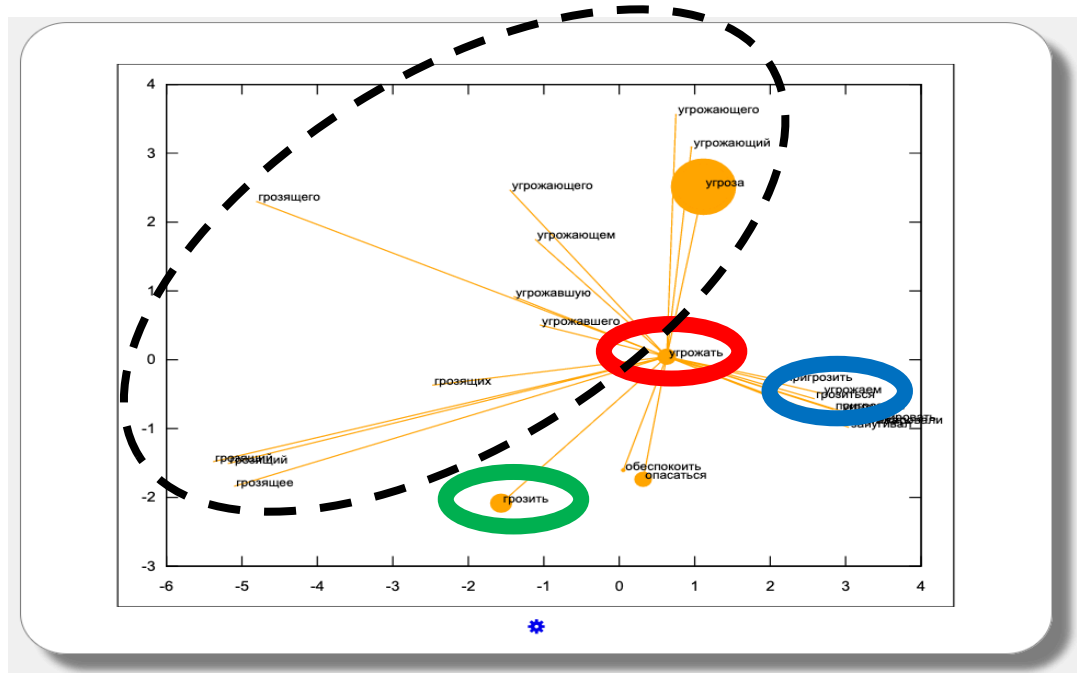
To summarize, our analysis of the semantic vectors from the Araneum Russicum Russicum Maius corpus strongly suggests that *grozit'* and *grozit'sja* are not complete synonyms. On the basis of corpus data, it should therefore be possible to pinpoint the differences between the two verbs – a task we turn to in the following sections.

3. Collocations

A useful function of the Araneum Russicum Russicum Maius corpus is to search for collocations, i.e., words that are likely to co-occur with *grozit'* and *grozit'sja*.[4] As we will see, the two verbs turn out to have different profiles when it comes to collocations. This lends further support to the observation that the two verbs are not perfect synonyms.

We searched for collocations of *grozit'* and *grozit'sja* with the specification that the distance between the verb and the other word be from +1 to -1 word. In this way, we identify the words immediately preceding and following the verbs under scrutiny. The corpus offers several ways of ranking the collocations. We chose the logDice option, which is useful for data from large corpora, since it does not take into account corpus size. The fifty most highly ranked collocations for *grozit'* and *grozit'sja* are listed in Table 2.

Grozit'	logDice	Grozit'sja	logDice
штраф	8.19887	взвинтить	6.56121
опасность	7.69859	выгнать	6.31525
лишение	7.27693	выселить	5.54810
наказание	6.91199	отомстить	5.51982
обернуться	6.88363	поджечь	5.20396
нарушитель	6.86265	отравить	5.12540
тюремный	6.84390	обрушиться	4.61792
гибель	6.45812	уволить	4.59462
перерасти	6.29104	вот-вот	4.43086
смертный	6.16966	сжечь	4.35690
неминуемый	6.04915	разорвать	4.26847
серьезный	5.85520	побить	4.19426
увольнение	5.78396	перерасти	4.18515
уголовный	5.66432	выписать	4.13041
пожизненный	5.63443	отнять	4.08039
пальчик	5.60695	выкинуть	4.01772
вымирание	5.58353	убить	3.97398
потеря	5.57361	забрать	3.84047
тюрьма	5.52686	наказать	3.47399
смертельный	5.51174	отобрать	3.41720
кулак	5.45412	подать	3.32377
беда	5.41713	заправка	2.88311
банкротство	5.28178	отменить	2.85241
арест	5.25167	лишить	2.73370
катастрофа	5.25131	бросить	2.63994
срыв	5.20595	разрушить	2.46576
обвал	5.13969	уйти	2.34544
затянуться	5.13802	сдать	2.24015

---

[4] We also checked the collocation function in CoCoCo (Collocations, Colligations, Corpora, https://cococo.cosyco.ru), but this tool did not return relevant results for *grozit'sja*, which is less frequent than *grozit'*.

санкциями	5.13228	посадить	2.15601
неприятность	5.11375	вывести	1.96016
летальный	5.01836	уничтожить	1.93767
дефолт	5.00699	закрыть	1.83435
дисквалификация	5.00468	превратиться	1.83407
административный	4.98508	приехать	1.53527
человечество	4.96115	отправить	1.50875
исчезновение	4.95020	позвонить	1.27813
обрушением	4.93257	чиновник	1.27485
чем	4.91842	яндекс	1.24431
промедление	4.88894	запустить	1.23064
осложнение	4.85271	опубликовать	1.10860
смерть	4.84071	муж	1.04212
вот-вот	4.81574	давно	0.98723
разорение	4.81207	царь	0.97841
превратиться	4.80409	парень	0.97348
разрушение	4.77729	ввести	0.91549
депортация	4.76885	передать	0.88388
водитель	4.76812	устроить	0.85994
обрушиться	4.71030	отдать	0.81448
участь	4.65922	повысить	0.78598
бесплодие	4.59036	оставить	0.65668

Table 2: The fifty most highly ranked collocations for grozit' and grozit'sja from the Araneum Russicum Russicum Maius corpus. Ranking according to logDice.

We would like to draw attention to two facets of the lists in Table 2. First, we see that the logDice values are generally higher for *grozit'* than for *grozit'sja*. This suggests that *grozit'sja* is more flexible with regard to the contexts it occurs in, while *grozit'* may have closer ties to its collocates. Second, the two lists are quite different, which shows that the two verbs typically combine with different words. The second point becomes even clearer if we classify the collocates with regard to their parts of speech. As shown in Table 3 and Figure 2, *grozit'* tends to combine with nouns and, to a lesser degree, adjectives, while *grozit'sja* typically cooccurs with verbs. The small category "other" in the table includes adverbs and pronouns. The differences are statistically significant with a large effect size.[5]

	grozit'	*grozit'sja*
Noun	35	5
Adjective	8	0
Verb	5	43
Other	2	2

Table 3: The fifty most highly ranked collocations for grozit' and grozit'sja sorted according to parts of speech

[5] We compared the numbers for nouns and verbs for *grozit'* and *grozit'sja*. Pearson's Chi-squared test with Yates' continuity correction (X-squared = 49.225, df = 1) returned a p-value = 2.282e-12. Cramer's V-value was calculated to 0.77, which indicates a large effect size.
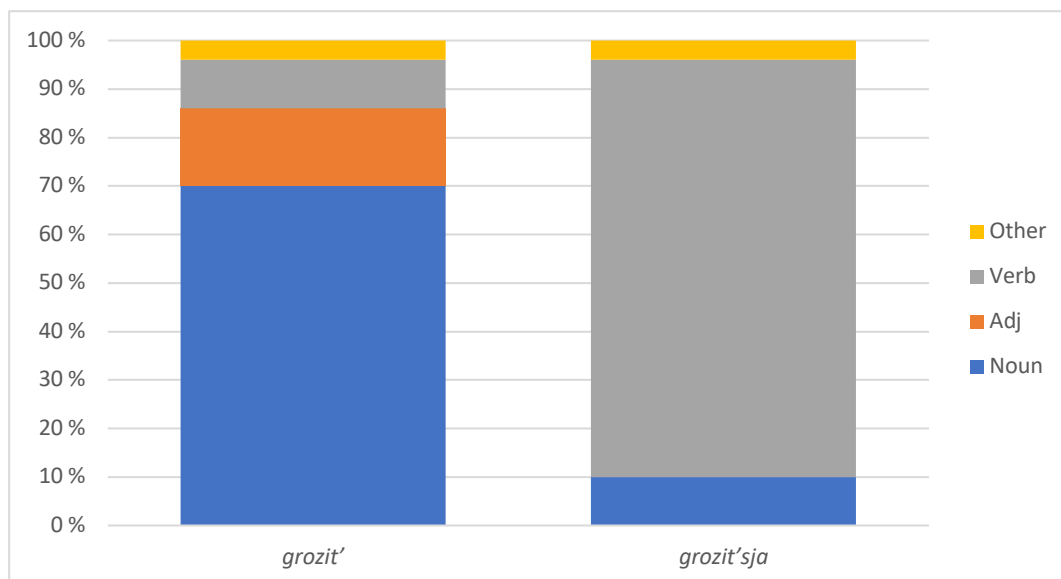
Figure 2: The fifty most highly ranked collocations for grozit' *and* grozit'sja *sorted according to parts of speech*

Our analysis of collocations in the Araneum Russicum Russicum Maius corpus brings us one step closer to pinpointing the differences between *grozit'* and *grozit'sja.* Knowing that the former prefers combinations with nouns, while the latter typically collocates with verbs, we can proceed to a more detailed analysis of the constructions, in which *grozit'* and *grozit'sja* occur.

4. Constructions: argument structure

In order to get a clearer picture of the constructions of *grozit'* and *grozit'sja*, we created a database with examples from the Russian National Corpus.[6] This corpus is smaller than the Araneum Russicum Russicum Maius corpus, but it is curated and provides considerable metadata for each example, thus facilitating in-depth analysis. Our data confirm the observations from the two previous sections that *grozit'* and *grozit'sja* show different behavior (see also Glovinskaja 2004a-b for discussion).

Our database was constructed as follows. We searched for both verbs in five time periods: 1800-1849, 1850-1899, 1900-1949, 1950-1999, 2000-present. For each period, we made a random sample of fifty examples for each verb. In order to avoid biased samples, we only included one example for each author. *Grozit'sja* is less frequent than *grozit'*, and for the 1800-1849 period we were only able to include 23 examples in the database. All in all, the database thus contains 473 examples – 250 for *grozit'* and 223 for *grozit'sja*. The examples were manually annotated for their syntactic constructions, as well as several other parameters, which we will come back to in later sections.

For the purposes of our analysis, we distinguish between five constructions. A frequent pattern is for the verbs to combine with a nominative subject, an argument in the dative and additional constituents. We refer to this construction as NomVDat+:[7]

(5)   I oni znali: u babuški Dženni **im ne grozjat** poučenija. (Danin 1969-1975)
       'And they knew: at grandmother Jenny's place they would not be threatened with any homilies.'

---

[6] We used the main subcorpus of the Russian National Corpus, which contains approximately 330 million words. Corpus searches were carried out on September 23, 2021.

[7] In the abbreviations for the constructions, V stands for the verb *grozit'* or *grozit'sja*.

(6) Poslednjuu tiradu on proiznës s bol'šoj siloj, budto **grozjas' komu-to**. (Elčaninov 1926)
'The last rant he pronounced very vigorously, as if he was threatening someone.'

We have quite a few examples where the verb cooccurs with a nominative subject and a complement in the instrumental, as well as additional constituents. We call this construction NomVInstr+:[8]

(7) Poètomu problemy so zdorov'em byli, a zabolevanie, načavšeesja 31 ijulja, v dal'nejšem **grozit ser'jëznymi osložnenijami**. (Lariševa 1997)
'So there were some health issues, and the illness that started 31 July, can have serious complications.'

(8) I ona daže **grozitsja pal'cem**. (Budišč'ev 1901)
'And she is even making threatening gestures with her finger.'

A combination of dative and instrumental complements are found in examples of the following type:

(9) Ja **ničem i nikomu ne grožu**. (Dombrovskij 1978)
'I am not threatening anyone with anything.'

(10) —U, zarazy, — **grozilas' derev'jam** babuška **suxon'kim kulakom**. (Olejnikov 2007)
'You bastards, the old woman threatened the trees with her dry fist.'

The previous examples involve nominal complements. However, another important construction, for which we use the label NomVInf, involves a nominative subject and an infinitive complement:

(11) Molčanov stal za nim, vynul špagu i **grozil izrubit'** ego, eželi on strusit. (Dolgorukov 1788)
'Molčanov stood behind him, he pulled out his sword and threatened that he would cut him in pieces if he should act like a coward.'

(12) Prišli medsëstry i skazali, čto on **grozitsja** ix vsex **poubivat'**. (Rudnickaja 2007)
'The nurses came and said that he was threatening to kill them all.'

It is not uncommon for *grozit'* and *grozit'sja* to occur with a subject in the nominative, but no complement. Examples of this NomV construction are:

(13) Ja ne **grožu**, ja ne vymogaju prošč'enija. (Brik 1956-1977)
'I am not threatening, neither am I begging for forgiveness.'

(14) On ne obižalsja, ne plakal, ne **grozilsja** [...]. (Nagibin 1994)
'He was not getting offended, did not cry or threaten […].'

Finally, we have a number of attestations where the verb cooccurs with a clausal complement in addition to the nominative subject. We refer to this construction as NomVClause:[9]

(15) Ešč'ë do ot''ezda on v razgovorax s drugimi **grozil, čto** Griboedovu èta šutka ne projdët darom. (Smirnov 1842-1866)

---

[8] Notice that the noun phrases in the instrumental can represent the potential consequence of the threatening situation as in (7) or the body part or weapon that is used in the relevant situation to threaten someone as in (8). Both types are attested for both verbs in our database.

[9] Notice that we also include examples where *grozit'* or *grozit'sja* are followed by direct speech in the NomVClause category.

'Even before he left, in conversations with the others he was threatening that this joke would not go without consequences for Griboedov.'

(16) Potom ja plakala i ona menja utešala, **grozilas', čto** sladkogo ne dast. (Katanjan 1998)
'Then I was crying, and she was comforting me, threatening that she would not give me any sweets.'

The distribution of these constructions in our database is summarized in Table 4 and Figure 3, where the category "other constructions" includes miscellaneous types, e.g., with prepositions or participles in oblique cases. As shown, the two verbs are attested in the same constructions, but nevertheless have different profiles. For *grozit'*, by far the most frequent construction is NomVDat+, which is rare for *grozit'sja*. Other frequent constructions for *grozit'* involve complements in the instrumental or the combination of instrumental and dative complements. By contrast, the most frequent option for *grozit'sja* is the NomVInf construction, which is much less frequently attested for *grozit'*. These results square with the findings from the Araneum Russicum Russicum Maius corpus presented in the previous section, insofar as *grozit'* typically combines with nominal arguments, whereas *grozit'sja* prefers an infinitive complement. The observed differences are statistically highly significant and show a large effect size.[10]

For the other constructions, the numbers are smaller and the differences less clear, but it is interesting to notice that the NomVInstr+ construction is more frequent for *grozit'* than for *grozit'sja*. Once again, we see that *grozit'* has the stronger affinity for nominal complements (here: an NP in the instrumental case).

	grozit'	*grozit'sja*
NomVDat+	70	5
NomVInstr+	44	25
NomVDatInstr	38	1
NomVInf	48	135
NomV	12	32
NomVSent	5	17
Other constructions	33	11
Total	250	223

Table 4: The distribution of constructions with grozit' *and* grozit'sja. Data from the Russian National Corpus.

---

[10] We compared the numbers for NomVDat+, NomVInstr+ and NomVDatInstr on the one hand with the numbers for NomVInf on the other. Pearson's Chi-squared test with Yates' continuity correction (X-squared = 116.95, df = 1) returned a p-value < 2.2e-16. Cramer's V-value is 0.6, which represents a large effect size.
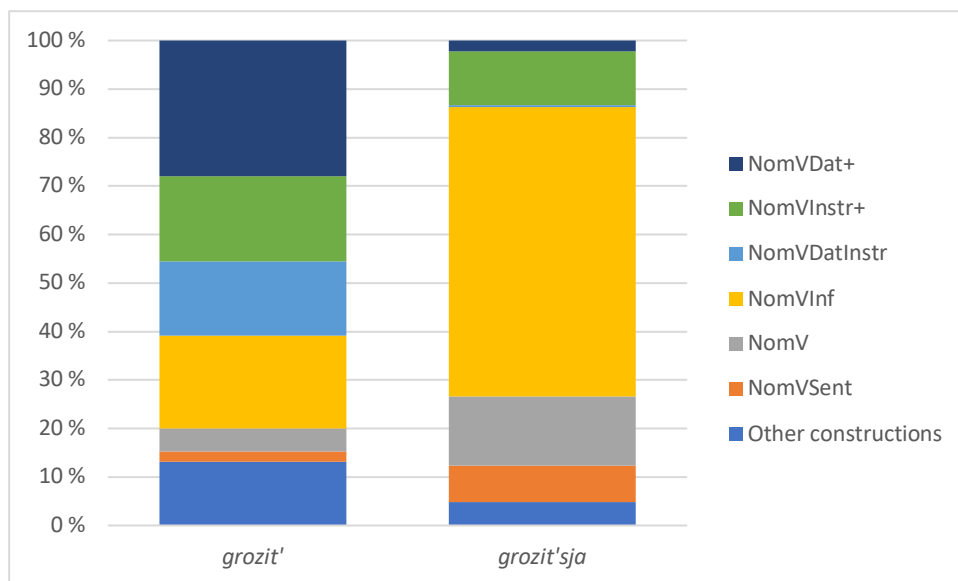
Figure 3: The distribution of constructions with grozit' *and* grozit'sja. *Data from the Russian National Corpus.*

To summarize, our analysis of data from the Russian National Corpus indicates that *grozit'* and *grozit'sja* are syntactically different, insofar as they tend to occur in different constructions. While *grozit'* typically takes a nominal complement in the dative or instrumental cases, *grozit'sja* is most frequently attested with an infinitive. We hasten to add that the observed differences are not categorical. Both verbs are attested in all the constructions we have explored in this section – but with very different frequencies.

5. Arguments: Persons vs. non-persons

A prototypical threat may be characterized as a situation where one person promises to do harm to another person, as in examples (3) and (4) cited in section 1. In other words, we are dealing with a relationship between two persons. In what follows, we show that this prototypical scenario is characteristic of *grozit'sja*, whereas *grozit'* has developed abstract meanings, following a grammaticalization path that is well known from other European languages (Heine and Miyashita 2007 and 2008, Narrog and Heine 2021).

Peškovskij (1956: 119) and Gerritsen (1990: 96) have mentioned that *grozit'sja* combines with subjects that refer to persons, while *grozit'* does not have such a restriction. In order to test this hypothesis against corpus data, we distinguish between two broad categories: "persons" and "non-persons", where the latter category includes both entities (concrete objects and abstract concepts) and events.[11] Here are relevant examples with non-persons:

(17) Za tjažkie prestuplenija ej **grozila smertnaja kazn'**. (Čertok 1999)
      'For her serious crimes a death penalty was threatening her.'
(18) Aprel' mesjac stojal v polovine, **dorogi grozilis'** sdelat'sja neproxodimymi. (Družinin 1857)
      'It was the middle of April, and the roads threatened to become impassable.'

Notice that it is not only the "threatener" that can be a "non-person". In the following examples, "threatenee" is not a person. In (19), Africa is threatened, and in (20) the "threatenee" is the sun:

---

[11] Notice that we classify words according to their literal meanings. Thus, *Afrika* in example (19) is classified as "non-person", even if it arguably may refer metonymically to the people in Africa.

(19) **Afrike grozit** isčeznovenie kofejnyx derev'ev. (*Znanie-sila*, 2013)
      'Africa is threatened by the extinction of coffee-trees.'
(20) **Grozitsja** napolzti na **solnce** oblako […]. (Lipatov 1984)
      'A cloud threatened to cover the sun.'

In Table 5, we summarize the situation for the four logical combinations of "person" and "non-person". The first two rows represent situations where the "threatener" is a person, while the two rows at the bottom involve situations where the "threatener" is not a person.[12]

	Grozit'	*Grozit'sja*
PersonToPerson	114	210
PersonToNon-Person	0	1
Non-personToPerson	83	8
Non-personToNon-Person	53	4

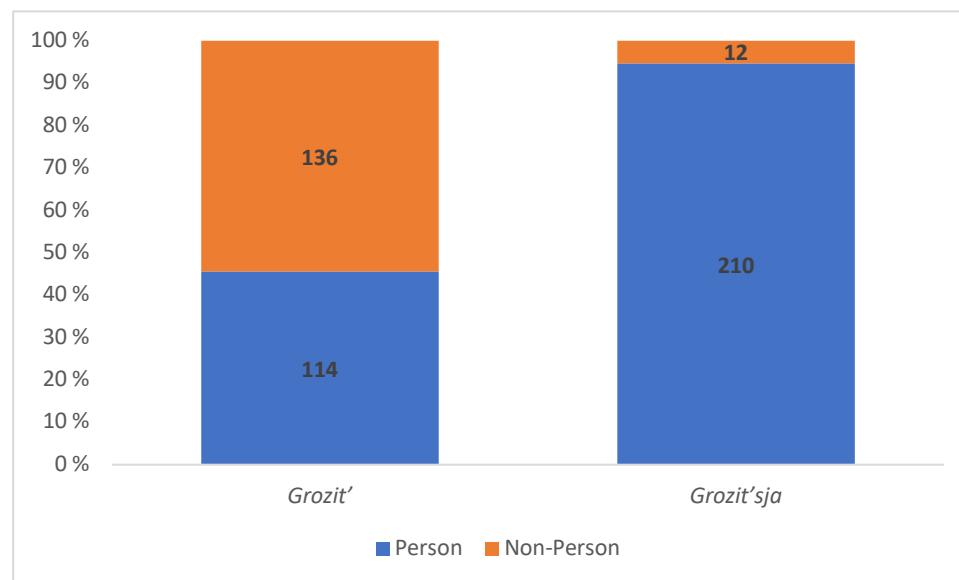Table 5: Persons and non-persons as arguments

Figure 4: Person vs. Non-Person as "threatener". The Numbers on the bars in the diagram are raw numbers.

The following observations can be made. First, we see that the prototypical threat (PersonToPerson) represents the most frequent option for both verbs. Second, the PersonToNon-Person is marginal. Third, the table shows that *grozit'* is well attested with a non-person as the "threatener", while this is not the case for *grozit'sja*. In other words, while *grozit'* is relatively evenly distributed between persons and non-persons as the "threatener", *grozit'sja* strongly prefer persons as arguments. Figure 4 visualizes the difference between persons and non-persons as the "threatener". The observed difference is statistically significant and has a large effect size.[13]

In section 1, we mentioned that verbs for 'threaten' have received considerable attention in studies of grammaticalization, since in many European languages the relevant verbs have

---

[12] Notice that the "threatenee" is not always explicitly marked (e.g., as a grammatical object) in the example sentences. In such cases, we have identified the "threatenee" on the basis of the wider context. The "threatenee" can be realized as noun phrases in different cases, as illustrated in (19) and (20).

[13] We compared examples with Person vs. Non-Person as the "threatener". Pearson's Chi-squared test with Yates' continuity correction (X-squared = 128.87, df = 1) returned a p-value < 2.2e-16. Cramer's V-value is 0.5, indicating a large effect size.

undergone grammaticalization from the prototypical scenario where one person threatens another person to more abstract meanings involving non-persons as arguments. The Russian data in Table 5 show a similar picture for Russian since non-persons are widespread. At the same time, Russian is different from the languages for which Heine and Miyashita (2007 and 2008) provide detailed analyses, because Russian has more than one morphologically related verb for 'threaten'. Table 5 suggests that it is mainly *grozit'* that follows the path of grammaticalization known from other European languages, whereas *grozit'sja* specializes on the prototypical situation where threats are relations between two persons.

A diachronic analysis lends further support to this conclusion. When we consider the development over time, we see that for *grozit'* the proportion of the PersonToPerson category has decreased over time. In the first half of the nineteenth century, about 60% of the examples with *grozit'* were of the PersonToPerson type, whereas in the beginning of the twenty-first century the corresponding number had decreased to approximately 20%. This difference is statistically significant with a moderate effect size.[14] For *grozit'sja*, on the other hand, the proportion of examples of the PersonToPerson type has been stably high over time. As shown in Table 5 and Figure 4, there are very few examples of the Non-Person type with *grozit'sja*, and this has not changed over time.

	1800-1850	1850-1899	1900-1950	1950-2000	2000-
grozit'	32 (64%)	22 (44%)	29 (58%)	18 (36%)	13 (26%)
grozit'sja	21 (91%)	48 (96%)	46 (92%)	49 (98%)	46 (92%)

Table 6:The proportion of the PersonToPerson category over time for grozit' *and* grozit'sja (raw numbers and per cent)
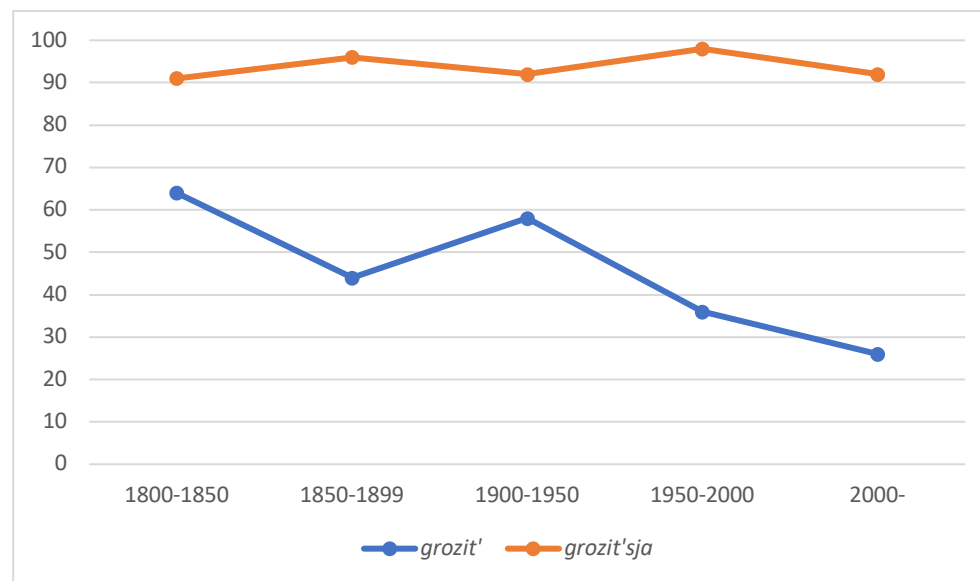
Figure 5: The proportion of the PersonToPerson category over time for grozit' *and* grozit'sja (per cent)

To summarize, our investigation shows that *grozit'sja* typically describes a relationship between two persons, whereas *grozit'* displays a more varied constructional profile. This difference has increased over time.

---

[14] We compared the numbers for *grozit'* in the first half of the 19th century and in the beginning of the 21st century. Pearson's Chi-squared test with Yates' continuity correction (X-squared = 16.552, df = 1) returned a p-value = 4.733e-05. Cramer's V-value was calculated to 0.4.

6. Verbal vs. gestural threats

After having focused on the arguments of the verbs, we now turn to the verbs themselves. In particular, we show that *grozit'sja* tends to involve verbal threats, whereas *grozit'* is more versatile.

Threats can be conveyed by means of words or by a physical gesture, a distinction that has been considered relevant for the choice between *grozit'* and *grozit'sja*. Gerritsen (1990: 96) suggests that *grozit' pal'cem* 'show a threatening gesture with a finger' is grammatical, whereas *grozit'sja* is not grammatical in this context. She furthermore observes that in a sentence like *Gurov opjat' pogrozil* 'Gurov threatened again' *grozit'* "typically refers to a threatening gesture" even if the relevant body part is not mentioned explicitly (Gerritsen 1990: 96). A similar view is expressed by Israeli (1997: 107), who suggests that *grozit'* is the preferred choice for non-verbal threats.

In order to test the relevance of verbal vs. gestural threats against corpus data, we annotated our database for three broad categories. "Gestural threats" involve examples where a gesture expressing the threat is explicitly mentioned in the context. The threatener can either use a body part or an object such as a weapon:

(21) […] kričal dvornik, **grozja** iz okna **ključami** […]. (Gončarov 1842)
       'The groundskeeper screamed, threatening with the keys from the window.'

(22) […] kriknul on, **grozjas' podnjatym arapnikom** na grafa. (Tolstoy 1867-1869)
       'He screamed, threatening the count with his raised whip.'[15]

Our category "Verbal threat" covers examples where the threat is conveyed by words and there is no evidence from the context that the threatening words are accompanied by a gesture:

(23) Nu, smotri! – **grozil** ej batjuška. (Zlatovratskij 1911)
       'Watch out, – the father threatened her.'

(24) Nu ja im! – **grozilsja** Šapošnikov. (Astaf'ev 1995)
       'I will show them! – threatened Šapošnikov.'

The situation is summarized in Table 7, where "other" refers to examples that do not belong to the two categories discussed above, such as sentences where the "threatener" is not a person. As shown, verbal threats are more frequently attested than gestural threats. This holds for both verbs, although the tendency is stronger for *grozit'sja* than for *grozit'*. Contrary to what Gerritsen (1990) proposed, gestural threats are attested for *grozit'sja*, which in our database combines with body parts (e.g., *kulak* 'fist', *palec* 'finger') and objects (e.g., *skalka* 'rolling pin', *arapnik* 'whip', *palaš* 'sword'). Israeli's suggestion that *grozit'* is the preferred choice for non-verbal threats is supported by our data. For gestural threats, the proportion of examples in our dataset is almost twice as large for *grozit'* compared to *grozit'sja*. The difference is even larger in the category "other", where we have more than ten times as many examples for *grozit'* as for *grozit'sja*. The large proportion of "other" threats for *grozit'* is related to the fact that *grozit'* often involves non-personal "threateners", as shown in the previous section.

---

[15] Notice that the "threatenee" is represented as a prepositional phrase with *na* 'on'. While the most common pattern is for the "threatenee" to be encoded as a noun phrase in the dative, we have six examples in our database with *na.* The most recent dated example in the Russian National Corpus is from 1937, which suggests that this pattern is somewhat archaic.

	Grozit'	*Grozit'sja*
Gestural threat	28	15
Verbal threat	81	195
Other	141	13

Table 7: The distribution of verbal and gestural threats for grozit' *and* grozit'sja

To summarize, our investigation of data from the Russian National Corpus supports the idea that the distinction between verbal and gestural threats is relevant for *grozit'* and *grozit'sja*. While *grozit'* is frequently used for both verbal, gestural and other threats, for *grozit'sja* verbal threats are the dominant type.

7. Consequences: serious or not?

The consequences of a threat may vary from very serious to not serious at all. This has been argued to be relevant for the choice between *grozit'* and *grozit'sja.* In the following, we present weak evidence that *grozit'* may be more compatible with serious consequences, but at the same time we show that it is difficult to test this hypothesis in a rigorous way.

Commenting on the difference between *grozit'* and *grozit'sja* and similar verb pairs, Israeli (1997: 107) argues that "the non-*sja* verb means an action that has impact." It is not straightforward to test this hypothesis against corpus data, since "an action that has impact" can be subjected to various interpretations. However, it seems that we would expect *grozit'* to involve more serious consequences than *grozit'sja*. Accordingly, we divided the examples into three broad categories: "very serious", "intermediate", and "non-serious". In the "very serious" category, we included threats involving death and complete destruction, while "non-serious" comprise ironic contexts where the threat is not seriously meant. The "intermediate" category contains all remaining examples, which cover a whole range of more or less serious threats.

	Grozit'	*Grozit'sja*
Very Serious	47	23
Intermediate	188	179
Non-serious	15	21

Table 8: Degree of seriousness for grozit' *and* grozit'sja

Table 8 suggests a tendency for "very serious" threats to favor *grozit'* over *grozit'sja.* The differences between "very serious" and "non-serious" are statistically significant with a small, but reportable effect size.[16] However, the numbers are small, and the assessment of the degree of seriousness is to some extent a subjective matter. It is furthermore difficult to control for the interaction with other factors. We conclude that more research is needed in order to better understand the relevance of serious vs. non-serious threats.

8. *Grozit'* vs. *grozit'sja* and the meaning and functions of -*sja*

What do our findings tell us about the meaning and functions of -*sja*? We will argue that -*sja* changes the argument structure and the meaning of the verb in a way that relates *grozit'sja* to the middle voice.

---

[16] We compared the numbers for "very serious" and "non-serious" threats. Pearson's Chi-squared test with Yates' continuity correction (X-squared = 5.3492, df = 1) returned a p-value = 0.02. Cramer's V-value 0.2.

Table 9 summarizes our findings. Recall from earlier sections that the results we report are statistical tendencies, rather than categorical rules. Our findings nevertheless show that *grozit'* and *grozit'sja* are semantically and syntactically distinct, although they display overlapping properties. Corpus data therefore clearly represent a valuable resource for the study of *-sja*, and a "local approach" studying individual verbs in detail has the potential to sharpen our understanding of *-sja*.

Topic	*Grozit'*	*Grozit'sja*	Section
Arguments – collocations	Nouns	Verbs	3
Arguments – syntax	NomVDat	NomVInf	4
Arguments – semantics	Person and non-person	Person	5
Arguments over time	Person decreases	Person stays high	5
Situation – type of threat	Non-verbal	Verbal	6
Consequences	More serious (?)	Less serious (?)	7

Table 9: Overview of findings – differences between grozit' *and* grozit'sja *as presented in sections 3 through 7*

Does *-sja* have an impact on the argument structure of the verb? We first consider the subject. In examples where *-sja* serves as a middle (reflexive) marker human or, at least, animate subject is required, since such sentences typically involve a human being carrying out a controlled action directed towards oneself. Good examples are "grooming verbs" such as *myt'sja* 'wash (oneself)' and *brit'sja* 'shave (oneself)'. The requirement of a human subject suggests that *grozit'sja* is closely related to examples where *-sja* is a middle (reflexive) marker.

With regard to objects, *-sja* typically entails decreased transitivity, since verbs with *-sja* normally do not combine with accusative objects.[17] We see subtle effects of *-sja* as a "detransitivizer" in many verbs. A case in point is the verb pair *brosat'–brosat'sja* 'throw', where *brosat'* is transitive and takes an object in the accusative, whereas *brosat'sja* combines with a complement in the instrumental case (Goto and Say 2009: 200):

(25) Kogda **brosaeš' kamni** v vodu, sledi za krugami, inače tvoë zanjatie budet bessmyslenno. (Dombrovskij 1978)
'When you throw rocks in water, pay attention to the rings, otherwise your activity becomes meaningless.'

(26) Živuščij v stekljannom dome ne dolžen **brosat'sja kamnjami**. (Valeeva 2002)
'A person who lives in a glass house should not throw rocks.'

Other examples where the accusative object of the non-*sja* verb is demoted to a complement in the instrumental case include *zadavat'sja voprosom* 'ask oneself a question', which corresponds to the synonymous *zadavat' sebe vopros* with a reflexive pronoun in the dative and a direct object in the accusative (Goto and Say 2009: 194).

A more radical effect of *-sja* as a detransitivizer is found in anticausatives such as *slomat'sja* 'break down' and autocausatives like *podnimat'sja* 'get up' (Goto and Say 2009: 194-195). These verbs do not take an object at all, as opposed to the corresponding transitive verbs *slomat'* 'break (something)' and *podnimat'* 'lift (something)' without *-sja*, which combine with direct objects in the accusative.

---

[17] A small group of verbs like *bojat'sja* 'fear' represent an exception to the general rule that *-sja* precludes objects in the accusative. For detailed analysis the reader is referred to Nesset and Kuznetsova (2015a-b). Notice that we follow Næss (2007) and Letučij (2014), who treat transitivity as a scalar phenomenon structured around a prototype with a direct object in the accusative case. Detailed discussion of this issue is beyond the scope of the present study.

Where does *grozit'sja* place itself in this picture? Does *-sja* serve as a "detransitivizer" involving object demotion? Providing a principled answer is not straightforward, since *grozit'* is not a transitive verb with an accusative object. As we have shown, *grozit'* typically combines with a dative and/or instrumental complement, while *grozit'sja* shows an affinity to infinitive complements. Which of these argument structures are most closely related to the transitive prototype with an accusative object? A possible criterion is the ability to undergo passivization. A sentence with a complement that can become the subject of a passive sentence is arguably closer to a prototypical transitive sentence than a sentence where passivization is impossible. Letučij (2014) observes that in general infinitive complements have fewer restrictions when it comes to passivization than do nominal complements in other cases than the accusative. If we take this observation seriously, we cannot say that *grozit'sja* is further removed from a prototypical transitive sentence than *grozit'*. At least, *grozit'sja* does not provide strong evidence for the detransitivizing effect of *-sja.*

Now that we have considered the arguments of the verbs, we must explore the effect of *-sja* on the meaning of the verb itself. We have shown that *grozit'sja* tends to involve verbal threats, possibly with less serious consequences than *grozit'.* It is not straightforward to see a connection to *-sja* in other verbs. This, on the other hand, may not come as a big surprise, since Goto and Say (2009: 188) observe that individual semantic idiosyncrasies are quite widespread among verbs with *-sja.* In this respect, *grozit'* and *grozit'sja* show similarities with many verb pairs such as *rešit'-rešit'sja* 'decide'. For instance, both *grozit'sja* and *rešit'sja* often combine with an infinitive complement that has a coreferential subject with the main verb. However, detailed comparison with such verb pairs is beyond the scope of the present study.

To summarize, our analysis shows that *grozit'sja* prefers persons as subject, a feature that relates the verb to verbs where *-sja* is a middle marker. At the same time, our analysis illustrates the ability of *-sja* to change the argument structure and meaning of a verb in somewhat idiosyncratic and unpredictable ways, to some extent dependent on the meaning of the base verb. In other words, simple labels like "middle voice" and "passive" are not sufficient for an adequate analysis. We need detailed studies of individual verbs to arrive at a deeper understanding of *-sja.* In short, we need to adopt a "local approach" to verbs with *-sja.*

9. Concluding remarks

In this study, we have provided a thorough analysis of *grozit'* and *grozit'sja*, using corpus data and methods of contemporary corpus linguistics. By way of conclusion, we would like to emphasize the following points. First, we have demonstrated that we are not dealing with complete synonyms, because the two verbs under scrutiny differ both syntactically and semantically. With regard to syntax, we have seen that *grozit'sja* is more likely to combine with human subjects and infinitive complements than is *grozit'*. In semantic terms, *grozit'sja* tends to involve verbal threats, while *grozit'* is often used about gestural threats. It is furthermore possible that *grozit'sja* implies less serious consequences than does *grozit'*, although we observe that it is difficult to test this hypothesis in a rigorous way.

A second finding concerns *-sja* – a descriptively and theoretically challenging morpheme in Russian. Our study testifies to the value of a "local approach" that considers individual verbs in detail. While categories like "middle voice" and "passive" are useful in the analysis of *-sja*, we also need detailed analyses of individual verbs in order to pinpoint all the idiosyncratic properties of verbs with *-sja.*

Third, our analysis has shown that Russian offers a welcome addition to the theoretical literature on the concept of "threat". In particular, Russian is interesting because it has more

than one morphologically related verb for 'threaten'. We have seen that Russian behaves like other European languages insofar as we find examples of abstract uses that deviate from the prototypical situation where one person promises to do harm to another person. However, our analysis shows that this primarily concerns *grozit'*, while *grozit'sja* is mostly used about prototypical threats.

A final point concerns methodology. Our analysis has illustrated the value of exploring semantic vectors and collocations in large internet corpora like the Araneum Russicum Russicum Maius. However, while these methods can give useful results, they can benefit from being supplemented with detailed analysis of concrete examples, preferably culled from curated corpora like the Russian National Corpus.

Although our analysis suffices to show that *grozit'* and *grozit'sja* are syntactically and semantically distinct, a more detailed analysis of a larger number of examples may shed more light on the differences between the two verbs – and on the meaning and functions of *-sja.* In particular, a detailed diachronic analysis of the two verbs would contribute relevant insights, as would a comparison to other Russian verbs for 'threaten', such as imperfective *ugrožat'* and perfective *pogrozit'* and *prigrozit'*. However, these and other issues are beyond the scope of the present study and must be left open for future research.

References

Cornillie, Bert. 2004. The shift from lexical to subjective readings of Spanish *prometer* 'to promise' and *amenazar* 'to threaten'. A corpus-based account. *Pragmatics* 14.1: 1–30.

Enger, Hans-Olav and Tore Nesset. 1998. The Value of Cognitive Grammar in Typological Studies: The Case of Norwegian and Russian Passive, Middle and Reflexive. *Nordic Journal of Linguistics* 22: 27–60.

Firth, John R. 1957. A synopsis of linguistic theory 1930-1955. In John R. Firth (Ed.): *Studies in Linguistic Analysis*. Oxford: Blackwell, 1–32.

Gerritsen, Nelleke. 1990. *Russian reflexive verbs: In search of unity in diversity*. Amsterdam and Atlanta: Rodopi.

Geniušienė, Emma. 1987. *The typology of reflexives*. Berlin, New York, Amsterdam: Mouton de Gruyter.

Glovinskaja, Marija Ja. 2004a. Grozit'. In Ju. D. Apresjan et al. *Novyj ob''jasnitel'nyj slovar' russkogo jazyka*. Moscow: Jazyki slavjanskoj kul'tury, 240-243.

Glovinskaja, Marija Ja. 2004b. Ugrožat'. In Ju. D. Apresjan et al. *Novyj ob''jasnitel'nyj slovar' russkogo jazyka*. Moscow: Jazyki slavjanskoj kul'tury, 1190-1194.

Goto, Ksenija V. and Sergey S. Say. 2009. Častotnye xarakteristiki klassov russkix refleksivnyx glagolov. In K.L. Kiseleva, V.A. Plungian, E.V. Raxilina and S.G Tatevosov (eds.): *Korpusnye issledovanija po russkoj grammatike.* Moscow: Probel-2000, 184–223.

Harris, Zellig S. 1954. Distributional Structure, *Word* 10.2-3: 146–162.

Heine, Bernd and Hiroyuki Miyashita 2007. The structure of a functional category: German *drohen*. *CogniTextes* 1: 1–23.

Heine, Bernd and Hiroyuki Miyashita. 2008. Accounting for a functional category: German *drohen* 'to threaten'. *Language Sciences* 30, 53–101

Israeli, Alina. 1997. *Semantics and pragmatics of the "reflexive" verbs in Russian*. Munich: Otto Sagner.

Joos, Martin. 1950. Description of language design. *Journal of the Acoustic Society of America* 22.6: 701–708.

Jurafsky, Daniel and James H. Martin. 2021. *Speech and Language Processing* (Ch. 6). Draft of December 29, 2021.

Kemmer, Suzanne. 1993. *The Middle Voice*. Amsterdam and Philadelphia: John Benjamins.

Knjazev Jurij P. 2007. *Grammatičeskaja semantika. Russkij jazyk v tipologičeskoj perspektive*. Moscow: Jazyki Slavjanskix Kultur.

Kyröläinen, Aki-Juhani. 2013. *Reflexive Space: A Constructionist Model of the Russian Reflexive Marker*. PhD Dissertation: University of Turku.

Langacker, Ronald W. 2000. *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.

Letučij, Aleksandr B. 2014. Perexodnosť. In *Russkaja korpusnaja grammatika:* http://rusgram.ru/%D0%9F%D0%B5%D1%80%D0%B5%D1%85%D0%BE%D0%B4%D0%BD%D0%BE%D1%81%D1%82%D1%8C.

Narrog, Heiko and Bernd Heine. 2021. *Grammaticalization*. Oxford: Oxford University Press.

Nesset, Tore and Julia Kuznetsova. 2015a. Constructions and language change: From genitive to accusative objects in Russian. *Diachronica* 32:3, 365-396.

Nesset, Tore and Julia Kuznetsova. 2015b. In Which Case Are Russians Afraid? *Bojat'sja* with Genitive and Accusative Objects. *Journal of Slavic Linguistics* 23.2, 255-283.

Næss, Åshild. 2007. *Prototypical Transitivity*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Peškovskij, Aleksandr M. 1956. *Russkij sintaksis v naučnom osveščenii* (izdanie sed'moe). Moscow: Gosudarstvennoe učebno-pedagogičeskoe isdatel'stvo ministerstva prosveščenija RSFSR.

Verhagen, Arie. 1995. Subjectification, syntax, and communication. In Dieter Stein and Susan Wright (Eds.): *Subjectivity and Subjectivisation: Linguistic Perspectives.* Cambridge: Cambridge University Press.