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

Tore Nesset and Anastasia Makarova

Abstract: This article explores the two verbs, *grozit'* and *grozit'sja*, which 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, although they share a number of properties, are semantically and syntactically distinct. 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 with regard to interactions between two persons, while *grozit'* has a wider range of uses. This tendency has become more pronounced 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.

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) voice

* We would like to express our gratitude to our colleagues in the CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) research group at UiT The Arctic University of Norway. Thanks to Sergey Say for discussing an earlier version of the paper with us, and to two anonymous reviewers and the editors of *JSL* for detailed and helpful comments.

in (1) and passive voice 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 **my-l-Ø-sja** pjat'
French king Louis XI wash-PST-SG-REFL five
raz-Ø v god-Ø.
time-GEN.PL in year-ACC.SG
'The French king Louis XI washed five times a year.'

(2) Kak ob″jasni-t' istori-ju pojavleni-ja na
how explain-INF history-ACC emergence-GEN on
territori-i Rossi-i nemetsk-ix kirx-Ø,
territory-LOC Russia-GEN German-GEN.PL church-GEN.PL
kotor-ye **stroi-l-i-s'** tevton-ami […]?
which-NOM.PL build-PST-PL-REFL teutons-INS.PL
'How can we explain the emergence of German churches on Russian territory, that were built by the Teutons […]?'

A note on terminology is necessary. Many researchers refer to examples like (1) as "reflexive" and verbs like *grozit'sja* as "reflexive verbs", 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'). Although the term "middle voice" has not been used so much in Russian and Slavic linguistics, we find it helpful since Russian has a grammatical distinction between *sebja* and *sja*, for which we can use the terms "reflexive" and "middle", respectively. This usage also comes with the advantage that it is in harmony with typological works on voice distinctions (e.g., Kemmer 1993).

Traditionally, *sja* has been described in terms of a list or, especially in cognitive and functionally-oriented linguistics, a network of related meanings or functions (see, for example, Geniušienė 1987; Enger and Nesset 1998; Goto and Say 2009; 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"

[1] All numbered examples are from the Russian National Corpus (www.ruscorpora.ru). For the convenience of the reader, the relevant verb or construction is boldfaced. Our database is available in TROLLing (The Tromsø Repository of Language and Linguistics); see Makarova and Nesset 2022.

descriptions of individual verbs (e.g., Israeli 1997; Knjazev 2007; 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, 2004b).[2] 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]

(3) Skol'ko raz-Ø Carevskij i Vevers
how.many time-GEN.PL Carevskij and Vevers
grozi-l-i mne sostavi-t' protokol o mo-ix
threaten-PST-PL I.DAT compile-INF protocol about my-LOC.PL
popytk-ax "diskreditirova-t' rukovodstv-o
attempt-LOC.PL discredit-INF leadership-ACC
obkom-a [...]".
regional.committee-GEN.SG
'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 často **grozi-l-i-s'** **Čebakov-u** **ujti**
you.PL often threaten-PST-PL-REFL Čebakov-DAT leave.INF
k svo-emu professor-u [...].
to own-DAT.SG.M professor-DAT.SG
'[Y]ou often threatened Čebakov to go to your professor […].'

[2] It is worth pointing out that Russian has a number of verb pairs with and without *sja*, which deserve closer analysis. Examples include *dymit' – dymit'sja* 'smoke', *kružit' – kružit'sja* 'spin', *rešit' – rešit'sja* 'decide', and *xvastat' – xvastat'sja* 'boast' (cf. Israeli 1997: 95–107; Gerritsen 1990: 95–97). For each of these pairs, it is difficult to pinpoint the exact semantic contribution of *sja*.

[3] Notice that "threatener" and "threatenee" do not have to be persons. For instance, in *Mne grozila smertnaja kazn'* (lit.) 'Death penalty threatened me', we analyze *smertnaja kazn'* 'death penalty' as the "threatener" (see also Section 5 below). We only analyze constituents that are overtly expressed in the examples. We would like to emphasize that "threatener" and "threatenee" are invariable semantic categories that can be realized as different syntactic functions (subject, object, etc.). While it would be interesting to carry out a systematic analysis of the relationship between "threatener" and "threatenee" on the one hand and syntactic functions on the other, such an analysis is beyond the scope of the present analysis.

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; 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 and functionally-oriented linguists have discussed the verbs for threatening in English, Dutch, and Spanish (cf., for example, Langacker 1999; Verhagen 1995; 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, 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 prefixed imperfective verb *ugrožat'*, as well as a number of prefixed perfective verbs, such as *prigrozit'*, which all can be translated as *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.* While 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).[4] 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 (2024). 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.[5] For each word one searches for, the corpus returns a list of the 25 most closely

[4] The Araneum Russicum Russicum Maius corpus is available at http://unesco.uniba.sk/.

[5] The function for assessing similarities among words can be found here: https://www.juls.savba.sk/sem%C3%A4/?lang=ru&kio=lemma&visualsel=gnuplot&topn=24&wpos=&wneg=. Our searches were carried out on 19 November 2021. A detailed discussion of the technical procedures behind the calculations of semantic vectors in the Araneum corpus is beyond the scope of the present study.

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'*. Both *grozit'* and *grozit'sja* are related to *ugrožat'*.

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

grozit'	**grozit'sja**	**ugrožat'**
0.000, grozit'	0.000, grozit'sja	0.000, ugrožat'
0.488, **ugrožat'**	0.295, grozilas'	0.470, ugrožajuščego
0.536, črevatyj	0.377, prigrozili	0.488, **grozit'**
0.562, grozjaščee	0.384, prigrozit'	0.519, ugroza
0.601, grozjaščij	0.406, prigrozila	0.520, ugrožajuščij
0.632, grozjaščego	0.510, grozjas'	0.559, ugrožavšego
0.635, grozjaščix	0.577, poobeščat'	0.609, **grozjaščix**
0.640, grozjaščij	0.611, obmateril	0.620, prigrozit'
0.644, povleč'	0.612, vygnat'	0.620, šantažirovat'
0.662, črevaty	0.613, otmestka	0.626, šantažirovali
0.666, grozivšij	0.613, voznamerit'sja	0.645, **grozit'sja**
0.668, obernut'sja	0.615, pripugnul	0.646, ugrožaem
0.683, grozivšaja	0.628, zasudjat	0.647, **grozjaščego**
0.687, karat'sja	0.630, zasudit	0.655, zapugival
0.691, neminuemyj	0.636, posmet'	0.663, zapugivat'
0.693, štraf	0.645, **ugrožat'**	0.671, ugrožaj
0.693, prigrozit'	0.649, šantažirovali	0.675, **grozjaščee**
0.696, sprovotsiruet	0.650, požaluetsja	0.677, obespokoit'
0.702, vleč'	0.651, naoral	0.684, **grozjaščij**
0.712, grozjaščem	0.662, vyšvyrnut'	0.684, **grozjaščij**

—continued on next page—

—continued—

grozit'	**grozit'sja**	**ugrožat'**
0.714, naneslo	0.665, nažalovalas'	0.685, opasat'sja
0.716, obespokoit'	0.666, zapugival	0.689, ugrožajuščem
0.716, prigrozili	0.666, šantažirovat'	0.693, ugrožajuščego
0.717, nakazyvat'sja	0.666, podgovorili	0.694, ugrožavšuju
0.717, grozivšej	0.667, nakažut	0.694, prigrozili

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*.[6] 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 50 most highly ranked collocations for *grozit'* and *grozit'sja* are listed in Table 2 on the following page.

[6] 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'*.

Table 2. The 50 most highly ranked collocations for *grozit'* and *grozit'sja* from the Araneum Russicum Russicum Maius corpus[7]

grozit'	**logDice**	***grozit'sja***	**logDice**
štraf	8.19887	vzvintit'	6.56121
opasnost'	7.69859	vygnat'	6.31525
lišenie	7.27693	vyselit'	5.54810
nakazanie	6.91199	otomstit'	5.51982
obernut'sja	6.88363	podžeč'	5.20396
narušitel'	6.86265	otravit'	5.12540
tjuremnyj	6.84390	obrušit'sja	4.61792
gibel'	6.45812	uvolit'	4.59462
pererasti	6.29104	vot-vot	4.43086
smertnyj	6.16966	sžeč'	4.35690
neminuemyj	6.04915	razorvat'	4.26847
ser'ëznyj	5.85520	pobit'	4.19426
uvol'nenie	5.78396	pererasti	4.18515
ugolovnyj	5.66432	vypisat'	4.13041
požiznennyj	5.63443	otnjat'	4.08039
pal'čik	5.60695	vykinut'	4.01772
vymiranie	5.58353	ubit'	3.97398
poterja	5.57361	zabrat'	3.84047
tjur'ma	5.52686	nakazat'	3.47399
smertel'nyj	5.51174	otobrat'	3.41720
kulak	5.45412	podat'	3.32377
beda	5.41713	zapravka	2.88311
bankrotstvo	5.28178	otmenit'	2.85241
arest	5.25167	lišit'	2.73370
katastrofa	5.25131	brosit'	2.63994
sryv	5.20595	razrušit'	2.46576
obval	5.13969	ujti	2.34544
zatjanut'sja	5.13802	sdat'	2.24015
sanktsijami	5.13228	posadit'	2.15601

—*continued on next page*—

[7] Collocations are ranked according to logDice. High numbers indicate a high likelihood for a word to occur next to *grozit'/grozit'sja*.

—*continued*— ***grozit'***	**logDice**	***grozit'sja***	**logDice**
neprijatnost'	5.11375	vyvesti	1.96016
letal'nyj	5.01836	uničtožit'	1.93767
defolt	5.00699	zakryt'	1.83435
diskvalifikatsija	5.00468	prevratit'sja	1.83407
administrativnyj	4.98508	priexat'	1.53527
čelovečestvo	4.96115	otpravit'	1.50875
isčeznovenie	4.95020	pozvonit'	1.27813
obrušeniem	4.93257	činovnik	1.27485
čem	4.91842	jandeks	1.24431
promedlenie	4.88894	zapustit'	1.23064
oslożnenie	4.85271	opublikovat'	1.10860
smert'	4.84071	muž	1.04212
vot-vot	4.81574	davno	0.98723
razorenie	4.81207	tsar'	0.97841
prevratit'sja	4.80409	paren'	0.97348
razrušenie	4.77729	vvesti	0.91549
deportatsija	4.76885	peredat'	0.88388
voditel'	4.76812	ustroit'	0.85994
obrušit'sja	4.71030	otdat'	0.81448
učast'	4.65922	povysit'	0.78598
besplodie	4.59036	ostavit'	0.65668

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 Figure 1 and Table 3, *grozit'* tends to combine with nouns and, to a lesser degree, adjectives, while *grozit'sja* typically co-occurs with verbs. The small category "other" in the table includes adverbs and pronouns. The differences are statistically significant with a large effect size.[8]

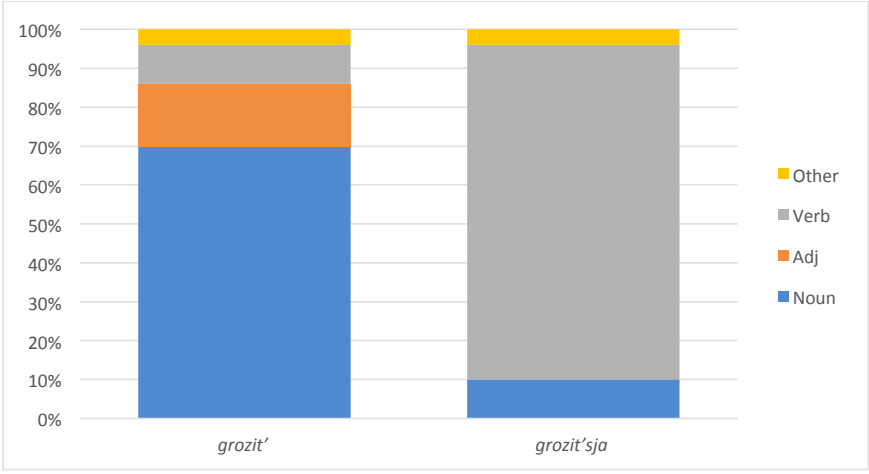

Figure 1. The 50 most highly ranked collocations for *grozit'* and *grozit'sja* sorted according to parts of speech

Table 3. The 50 most highly ranked collocations for *grozit'* and *grozit'sja* sorted according to parts of speech

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

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,

[8] 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.

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.[9] 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, 2004b 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 50 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+":[10]

(5) I oni zna-l-i: u babuški Dženni
and they know-PST-PL at grandmother Jenny
im **ne** **groz-jat** poučeni-ja.
they.DAT not threaten-3PL homily-NOM.PL
'And they knew: at grandmother Jenny's place they would not be threatened with any homilies.'

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

[10] In the abbreviations for the constructions, "V" stands for the verb *grozit'* or *grozit'sja*. The + sign indicates the possibility of additional arguments in the construction. Notice that the order of constituents has not been taken into consideration. In (5) and (6), for instance, the dative argument occurs in different positions, but we analyze both examples as the same construction.

(6) Posledn-juju tirad-u on proiznës-Ø s
last-ACC.SG.F rant-ACC.SG he pronounce-PST.SG.M with
bol'š-oj sil-oj, budto **groz-ja-s'** **komu-to.**
big-INS.F force-INS.SG as.if threaten-CVB-REFL someone.DAT
'The last rant he pronounced very vigorously, as if he was threatening someone.'

We have quite a few examples where the verb co-occurs with a nominative subject and a complement in the instrumental, as well as additional constituents. We call this construction "NomVInstr+":[11]

(7) Poètomu problem-y so zdorov'-em by-l-i, a
so problem-PL with health-INS.SG be-PST-PL and
zabolevanie, nača-vš-ee-sja 31 ijul-ja, v
illness start-PTCP-N.SG-REFL 31 july-GEN.SG in
dal'nejš-em **groz-it** **ser'jëzn-ymi**
future-LOC threaten-3SG serious-INS.PL
osložnenij-ami.
complication-INS.PL
'So there were some health issues, and the illness that started 31 July, can have serious complications.'

(8) I ona daže **groz-it-sja** **pal'c-em**.
and she even threaten-3SG-REFL finger-INS.SG
'And she is even making threatening gestures with her finger.'

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

(9) Ja **nič-em** **i** **nik-omu** **ne**
I nothing-INS.SG and no.one-DAT.SG not
grož-u.
threaten-PRS.1SG
'I am not threatening anyone with anything.'

[11] 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.

(10) — U, zaraz-y, — **grozi-l-a-s'** **derev'-jam**
oh bastard-NOM.PL threaten-PST-F-REFL tree-DAT.PL

babuška **suxon'k-im** **kulak-om**.
grandmother dry-M.INS.SG fist-INS.SG

'Oh, 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 sta-l-Ø za nim,
Molčanov stand-PST-M.SG behind he.INS

vynu-l-Ø špag-u i **grozi-l-Ø** **izrubi-t'**
take out-PST-M.SG sword-ACC and threaten-PST-M.SG cut-INF

ego, eželi on strus-it.
he.ACC if he act like a coward-FUT.3SG

'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) Priš-l-i medsëstr-y i skaza-l-i, čto
come-PST-PL nurse-NOM.PL and say-PST-PL that

on **groz-it-sja** ix vs-ex **poubiva-t'**.
he threaten-PRS.3SG-REFL they.ACC all-ACC.PL kill-INF

'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 demonstrated in the following examples:

(13) Ja ne **grož-u**, ja ne
I not threaten-PRS.1SG I not

vymoga-ju prošč'enij-a.
extort-PRS.1SG forgiveness-GEN.SG

'I am not threatening, neither am I extorting forgiveness.'

(14) On ne obiža-l-Ø-sja, ne plaka-l-Ø, ne
he not offend-PST-M-REFL not cry-PST-M not

grozi-l-Ø-sja [...].
threaten-PST-M-REFL

'He was not getting offended, did not cry or threaten […].

Finally, we have a number of attestations where the verb co-occurs with a clausal complement in addition to the nominative subject. We refer to this construction as "NomVClause".[12]

(15) Ešč″ë do ot″ezd-a on v
already before departure-GEN he in

razgovor-ax s drug-imi **grozi-l-Ø**, **čto**
conversation-LOC.PL with other-INS.PL threaten-PST-M that

Griboedov-u èt-a šutka ne projd-ët
Griboedov-DAT this-F.SG joke not pass-FUT.3SG

darom.
for.free

'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 plaka-l-a i ona menja
then I cry-PST-F and she I.ACC

uteša-l-a, **grozi-l-a-s'**, **čto** sladk-ogo ne
comfort-PST-F threaten-PST-F-REFL that sweet-GEN.SG not

da-st.
give-FUT.3SG

'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 Figure 2 and Table 4 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'*

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

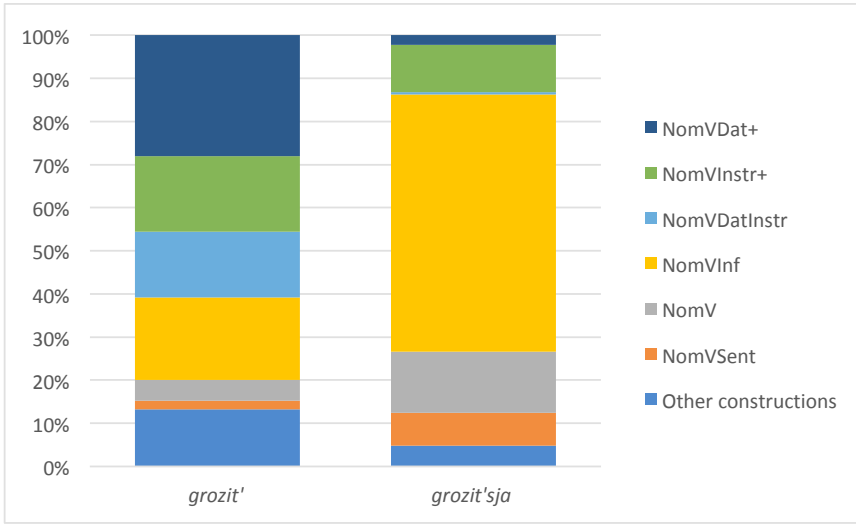

Figure 2. The distribution of constructions with *grozit'* and *grozit'sja* (data from the Russian National Corpus)

Table 4. The distribution of constructions with *grozit'* and *grozit'sja* (data from the Russian National Corpus)

	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

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.[13]

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).

Before we conclude, two methodological points deserve mention. First, we have focused on the constructions that are attested in the data. Here, we follow the usage-based approach of cognitive linguistics, where generalizations are assumed to be based on the patterns in actual language usage (Langacker 1991: 261–88 and 1999: 91–146). This methodology allows us to establish typical patterns, but we are not in a position to identify all possible constructions. Furthermore, we cannot identify which constructions are impossible.[14]

A second methodological point concerns the level of analysis. For the purposes of our study, we have characterized the arguments in terms of case and distinguished between nominal and clausal complements. It would be possible to create a more detailed analysis by adding, for instance, semantic roles. While this would have added another dimension to the analysis, it would yield a large number of small categories, on the basis of which no statistically robust generalizations could be made. We have therefore not added further semantic layers to our classification of constructions.

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, a fact we will return to in Section 8. 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.

[13] 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's 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.

[14] An anonymous reviewer points out that a dative argument is hardly compatible with a complement clause. According to him/her, examples like *On grozil emu vygnat'* 'He threatened to chase him away' without an explicit object in the embedded clause are completely unacceptable. We share the intuitions of the reviewer and agree that an investigation of such restrictions might be fruitful. However, in order to test the hypothesis of the reviewer properly, we would need an extensive survey with a large number of native speakers. Such an investigation is beyond the scope of the present study.

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, 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.[15] Here are relevant examples with non-persons:

(17) Za tjažk-ie prestupleni-ja ej
for serious-ACC.PL crime-ACC.PL she.DAT

grozi-l-a **smertnaja** **kazn'.**
threaten-PST-F death.ADJ penalty

'For her serious crimes a death penalty was threatening her.'

(18) Aprel' mesjac stoja-l-Ø v polovin-e, **dorog-i**
April month stand-PST-M in middle-LOC road-NOM.PL

grozi-l-i-s' sdela-t'-sja neproxodim-ymi.
threaten-PST-PL-REFL become-INF-REFL impassible-INS.PL

'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, the "threatenee" is not a person. In (19), Africa is threatened, and in (20), the "threatenee" is the sun:

(19) **Afrik-e** **groz-it** isčeznovenie kofejn-yx
Africa-DAT threaten-PRS.3SG extinction coffee-GEN.PL

derev'-ev.
tree-GEN.PL

'Africa is threatened by the extinction of coffee-trees.'

[15] 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.

(20) **Groz-it-sja** napolz-ti na **solnc-e** oblako […].
threaten-PRS.3SG-REFL cover-INF on sun-ACC cloud
'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.[16]

Table 5. Persons and non-persons as arguments

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

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 3 visualizes the difference between persons and non-persons as the "threatener". The observed difference is statistically significant and has a large effect size.[17]

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 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

[16] 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).

[17] We compared examples with Person vs. Non-Person as the "threatener". Pearson's Chi-squared test with Yates's 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.

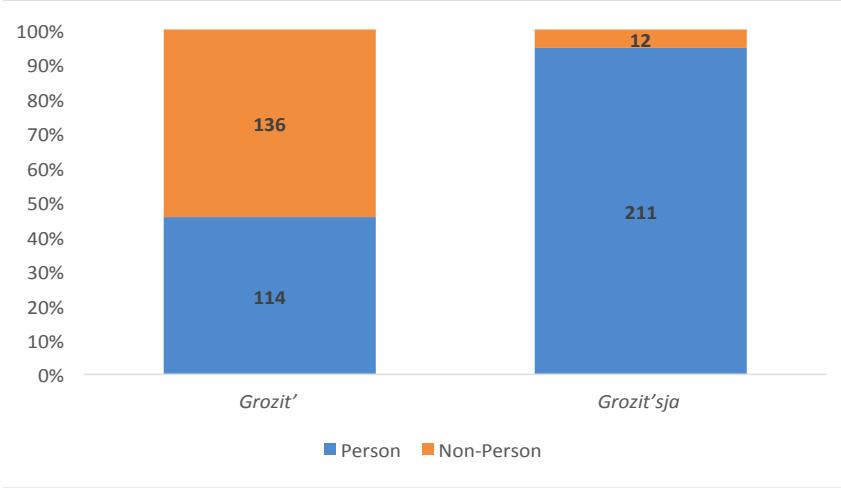

Figure 3. Person vs. Non-Person as "threatener"
(numbers on the bars in the diagram are raw numbers)

widespread. At the same time, Russian is different from the languages for which Heine and Miyashita (2007, 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 19th century, about 60% of the examples with *grozit'* were of the PersonToPerson type, whereas in the beginning of the 21st century, the corresponding number had decreased to approximately 20%. This difference is statistically significant with a moderate effect size.[18] 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 3, there are very few examples of the Non-Person type with *grozit'sja*, and this has not changed over time. The historical development is shown in Figure 4 and Table 6 on the following page.

[18] 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's continuity correction (X-squared = 16.552, df = 1) returned a *p*-value = 4.733e-05. Cramer's V-value was calculated to 0.4.

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.

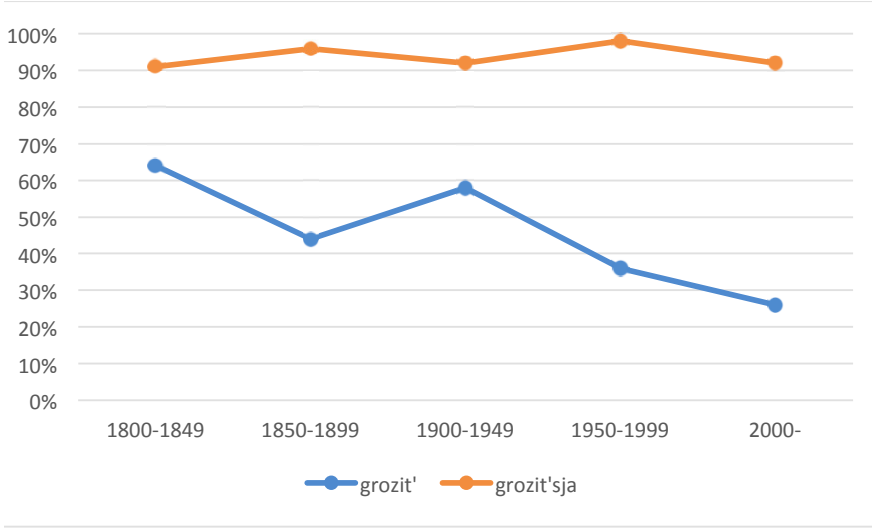

Figure 4. The proportion of the PersonToPerson category over time for *grozit'* and *grozit'sja* (per cent).

Table 6. The proportion of the PersonToPerson category over time for *grozit'* and *grozit'sja* (raw numbers and per cent)[19]

	1800–1849	**1850–1899**	**1900–1949**	**1950–1999**	**2000–**
grozit'	32 (64%)	22 (44%)	29 (58%)	18 (36%)	13 (26%)
grozit'sja	21 (91%)	48 (96%)	46 (92%)	49 (98%)	46 (92%)

6. Verbal vs. Gestural Threats

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

[19] For each cell in the table, the total is 50 examples. The only exception is the period 1800–1849 for *grozit'sja*, where we have only 23 examples in our database.

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) […] kriča-l-Ø dvornik, **groz-ja** iz
scream-PST-M groundskeeper threaten-CVB from
okn-a **ključ-ami […].**
window-GEN.SG keys-INS.PL
'The groundskeeper screamed, threatening with the keys from the window.'

(22) […] kriknu-l-Ø on, **groz-ja-s'** **podnja-t-ym**
[…] scream-PST-M he threaten-CVB-REFL lift-PST.PASS.PTCP-INS.SG
arapnik-om na graf-a.
whip-INS.SG on count-ACC
'He screamed, threatening the count with his raised whip.'[20]

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, smotr-i! **– grozi-l-Ø** ej batjuška.
so watch-IMP.SG threaten-PST-M she.DAT father
'Watch out, – the father threatened her.'

(24) Nu ja im! **– grozi-l-Ø-sja** Šapošnikov.
so I they.DAT threaten-PST-M-REFL Šapošnikov
'I will (show) them! – threatened Šapošnikov.'

[20] 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.

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.

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

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

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 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". This is an interesting observation that deserves discussion, although it is far from 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*. The question is: what counts as "serious consequences"? Since we are dealing with a scalar phenomenon which is difficult to quantify, it is not easy to avoid subjectivity completely. In order to reduce the level of subjectivity, we decided to focus on the end points of the scale. 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, which are events where the serious effect on the "threatenee" is uncontroversial:

(25) Naruši-vš-emu zapret-Ø **grozi-l-o**
break-PST.ACT.PTCP-DAT.M prohibition-ACC threaten-PST-N

proklat'e i skor-aja smert'-Ø.
curse and soon-ADJ.F.NOM death-NOM

'Those who did not respect the prohibition were threatened with a curse and an imminent death.'

(26) Zaduši-t' vsë **grozi-l-Ø-sja,** a potom
strangle-INF all threaten-PST-M-REFL and later

i zastreli-l-Ø.
and shoot-PST-M.SG

'He was threatening to strangle her, and then also shot her dead.'

At the other end of the scale, our category "non-serious" comprises ironic contexts where the threat is not seriously meant. In (27), it is clearly not a real threat that a theater would show "unprecedented decorative installations". Example (28) is about a party where the invited person promises to come. The use of 'threat' instead of 'promise' is ironic.

(27) Teatr **groz-it** pokaza-t' Parіž-u
theater threaten-PRS.3SG show-INF Paris-DAT

nevedom-ye dekorativn-ye ustanovk-i.
unprecedented-ACC.PL decorative-ACC.PL installation-ACC.PL

'The theater threatened to show Paris unprecedented decorative installations.'

(28) V 10.00 **grozi-l-Ø-sja** by-t' kak štyk.
at 10.00 threaten-PST-M-REFL be-INF as spit

'At 10.00 he threatened to be there guaranteed.'

The "intermediate" category contains all remaining examples, which cover a whole range of more or less serious threats.

It is instructive to leave the intermediate category aside and compare numbers of the extreme categories "very serious" and "non-serious", which involve the lowest degree of subjectivity in the classification. Table 8 suggests a tendency for "very serious" threats to favor *grozit'* over *grozit'sja*.

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

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

The differences between "very serious" and "non-serious" are statistically significant with a small, but reportable, effect size.[21] However, we are not dealing with large numbers, and as mentioned, 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.

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*.

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

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

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

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, a 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.[22] 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; see also Jakobson 1984: 79–80):

[22] A small group of verbs like *bojat'sja* 'fear' represent an exception to the general rule that *sja* precludes objects in the accusative. For a detailed analysis, the reader is referred to Nesset and Kuznetsova 2015a, 2015b. 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 (see also Hopper and Thompson 1980 and Chvany 1990). Detailed discussion of this issue is beyond the scope of the present study.

(29) Kogda **brosa-eš'** **kamn-i** v vod-u,
when throw-PRS.2SG stone-ACC.PL in water-ACC.SG
sled-i za krug-ami, inače tvoë zanjatie
watch-IMP.2SG after circle-INS.PL otherwise your activity
bud-et bessmyslenno.
be-FUT.3SG meaningless
'When you throw stones into water, watch the circles, otherwise your activity becomes meaningless.'

(30) Živu-šč-ij v stekljann-om dom-e
live-PRS.ACT.PTCP-NOM.M in glass-LOC house-LOC
ne dolžen-Ø **brosa-t'-sja** **kamn-jami.**
not shall-SG.M throw-INF-REFL stone-INS.PL
'A person who lives in a glass house should not throw stones.'

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: 194–95). 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.

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 voice 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. We have seen that *grozit'sja* is more likely to combine with human subjects and infinitive complements than is *grozit'*. Moreover, *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.

Sources

Araneum Russicum Russicum Maius. (2015) Russian web corpus. Available at: http://unesco.uniba.sk/.

Nacional'nyj korpus russkogo jazyka. (2003–2023) Available at: https://ruscorpora.ru/.

References

Chvany, Catherine V. (1990) "A continuum of lexical transitivity: Slightly-transitive verbs". Martina Björklund, Helena Lundberg, and Janina Orlov, eds. *Carina Amicorum*. Åbo: Åbo Academy Press, 51–63.

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". John R. Firth, ed. *Studies in linguistic analysis*. Oxford: Blackwell, 1–32.

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

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

Glovinskaja, Marija Ja. (2004a) "Grozit'". Ju. D. Apresjan et al., eds. *Novyj ob"jasnitel'nyj slovar' sinonimov russkogo jazyka*. Moscow: Jazyki slavjanskoj kul'tury, 240–43.

————. (2004b) "Ugrožat'". Ju. D. Apresjan et al., eds. *Novyj ob''jasnitel'nyj slovar' russkogo jazyka*. Moscow: Jazyki slavjanskoj kul'tury, 1190–94.

Goto, Ksenija V. and Sergey S. Say. (2009) "Častotnye xarakteristiki klassov russkix refleksivnyx glagolov". Ksenija 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): 146–62.

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

————. (2008) "Accounting for a functional category: German *drohen* 'to threaten'". *Language sciences* 30: 53–101.

Hopper, Paul J. and Sandra A. Thompson. (1980) "Transitivity in grammar and discourse". *Language* 56(2): 251–99.

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

Jakobson, Roman O. (1984) "Contribution to the general theory of case: General meanings of the Russian cases". Linda R. Waugh and Morris Halle, eds. *Russian and Slavic grammar: Studies 1931–1981*. Berlin/New York/Amsterdam: Mouton, 59–104.

Joos, Martin. (1950) "Description of language design". *Journal of the Acoustic Society of America* 22(6): 701–08.

Jurafsky, Daniel and James H. Martin. (2021) *Speech and language processing*. See esp. chap. 6, "Vector Semantics and Embeddings". Draft of 3 February 2024. Available at: https://web.stanford.edu/~jurafsky/slp3/.

Kemmer, Suzanne. (1993) *The middle voice*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

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*. Ph.D. dissertation, University of Turku.

Langacker, Ronald W. (1991) *Concept, image, and symbol*. Berlin: Mouton de Gruyter.

————. (1999) *Grammar and conceptualization*. Berlin: Mouton de Gruyter.

Letučij, Aleksandr B. (2014) "Perexodnost'". *Russkaja korpusnaja grammatika*. Available at: 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.

Makarova, Anastasia and Tore Nesset. (2024) "Replication Data for: Threatening in Russian with or without *-sja: grozit'* vs. *grozit'sja"*. Available at: https://doi.org/10.18710/J3UCVC. DataverseNO, V1.

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): 365–96.

————. (2015b) "In which case are Russians afraid? *Bojat'sja* with genitive and accusative objects". *Journal of Slavic linguistics* 23(2): 255–83.

Næss, Åshild. (2007) *Prototypical transitivity*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Peškovskij, Aleksandr M. (1956) *Russkij sintaksis v naučnom osveščenii.* 7th edition. Moscow: Gosudarstvennoe učebno-pedagogičeskoe isdatel'stvo ministerstva prosveščenija RSFSR.

Verhagen, Arie. (1995) "Subjectification, syntax, and communication". Dieter Stein and Susan Wright, eds. *Subjectivity and subjectivisation: Linguistic perspectives.* Cambridge, UK: Cambridge University Press.

Tore Nesset
Department of Language and Culture
UiT The Arctic University of Norway
Tromsø, Norway
tore.nesset@uit.no

Anastasia Makarova
Department of Modern Languages
Uppsala University
Uppsala, Sweden
anastasia.makarova@moderna.uu.se