The Facets of Sight: Slavic Verbs of Visual Perception in a Parallel Corpus

Maria Ovsjannikova

Draft for the discussion on the Colloquium, 27. 11.2024

1. Introduction

Considerable advances have recently been made in the study of lexical semantics and colexification patterns of perception verbs (San Roque et al. 2018; Georgakopoulos et al. 2021; Norcliffe, Majid 2024, among others). Still, due to their large cross-linguistic scale and the nature of the data, they are mostly restricted to basic perception verbs, such as *see* and *listen* in English. This study takes a close look at verbs of sight in three Slavic languages: Bulgarian, Polish, and Russian, which represent each of the three traditionally identified branches, i.e., South, West, and East Slavic, respectively.

Sight has been argued to be cross-linguistically the most prominent sense modality in terms of its textual frequency, lexical expression across languages, and the ability to develop non-perceptual meanings, see the hierarchy proposed by Viberg (1984, 2001), as well as Sweetser (1990: 39-40), San Roque et al. (2015; 2018). The present study analyses the correspondences between a wide range of verbs of sight across the three languages using the data of the Intercorp Parallel Corpus (Rosen et al. 2022; Rosen 2023). Example (1) shows a context where the basic imperfective verbs of sight correspond to each other in the three languages: B *vizhdam*, P *widzieć* and R *videt'* 'see' and B *gledam*, P *patrzeć* and R *smotret'* 'look'[1].

(1)   *Kade e Kolya? — izvika toy, kato **gledashe** Kolya, bez da go **vizhda**.*
          *Gdzie jest Kola? - krzyknął, **patrząc** na Kolę i nie widząc go.*
          *Gde Kolja? – vskrichal on, **smotrja** na Kolju i ne **vidja** ego.*
          'Where is Kolja? he asked looking at Kolja and not seeing him.' (author's translation)

However, there are many sight verbs beyond the basic ones and various cases of correspondence between verbs can be found in parallel contexts. For instance, in (2), there are three events expressed by sight verbs, and in none of them basic sight verbs are used consistently across the three languages. Explain the example in the text? Indicate which of the verbs are basic and which are not?

(2)   *<…> otnachalo toy go **izgleda** ravnodushno, setne se nadigna i raztarka ochi, ala kogato **pogledna** otnovo, veche ne go **vidya**.*
          *<…> vnachale on **smotrel** nevnimatel'no, no potom sel i proter glaza, no kogda on **vzgljanul** snova, nichego ne bylo **vidno**.*

---

*<…> w pierwszej chwili przyglądał się z roztargnieniem, potem wyprostował się i przetarł oczy kułakiem, lecz kiedy **spojrzał** znów w to samo miejsce – nic tam już nie **zobaczył**.*

'At first he stared at it listlessly, then he sat up and rubbed his eyes; but when he looked again he could not see it any more.'

In this study, the frequencies of correspondences between verbs in parallel contexts are used to establish the similarity in their meanings and usage. Then the similarity relations between verbs are investigated and the parameters relevant for the structuring of the domain of sight at least across these three languages are identified.

By taking a wider range of verbs into analysis, this study goes beyond the widespread paradigmatic view of the domain of perception, as elaborated by Viberg (1984; 2001). One of the structuring parameters of this paradigm is sense modality, which traditionally includes sight, hearing, smell, taste, and touch. For each of these sense modalities, the perceiver may either consciously direct their attention to an object, as exemplified by the English verbs *look* and *listen*, or perceive it without necessarily intending to do so, as in the case of the verbs *see* and *hear*. This distinction has been discussed under different terms, including activity vs. experience (Viberg 2001), active vs. passive perception (Nesset et al. 2008), opportunistic vs. explorative perception (Wälchli 2016). In this paper, the terms activity vs. experience will be used to denote this contrast. Along with the verbs where the perceiver is in the subject position, for each sense modality there may also exist verbs with the perceived object in the subject position, such as English *sound*. The latter type of verbs will not be considered in the present study. The paradigm of perception verbs for Russian is discussed by Padučeva (2004: 204) and Divjak (2015).

The paradigmatic approach to the study of perception was criticized by Wälchli (2016: 65), whose research on perception is an important point of reference for this study. In particular, Wälchli (2016: 65) suggests that the focus on the basic verbs of perception results in overlooking of the diversity of perception verb types and represents the semantic distinctions in the perception domain as more discrete and uniform across languages than they are in reality, see also Wälchli (2016: 63-64) on bias toward discrete features. The alternative approach he proposes looks at the distribution of verbs across parallel contexts to find the semantic groups of uses that arise from the data. Although different in terms of the specific methodology employed, the present study draws on his ideas of data-driven research of perception that looks at a wider range of perception types.

Apart from the structure of the sight domain, this paper addresses the issue of semantic generality as opposed to semantic specificity and the ways to quantitatively assess them. These notions can be understood differently depending on the domain of inquiry and is often left without an explicit definition. One of the possible definitions is that semantically general verbs "provide the central means by which humans are able to describe their experiences via a linguistic code" (Theakston et al. 2004:

62). In grammaticalization, semantic generality is typically understood as "abstract, schematic word meaning" (Hilpert & Correia Saavedra 2017: 370), a characteristic of grammatical items as opposed to lexical. Under any of these views, semantic generality is expected to be positively associated with frequency. The studies on verb acquisition by Theakston et al. (2004) and on word dispersion in texts by Hilpert & Correia Saavedra (2017) aim to disentangle the effects of semantic generality and frequency. In the present study, frequency and semantic generality are also treated as separate properties of verbs, and two methods for assessing different aspects of the semantic generality of sight verbs are proposed.

The paper is structured as follows. Section 2 gives an overview of the semantic distinctions and meaning extensions that can be expected to play a role in the structuring of sight verbs or may affect the degree of their mutual correspondence. In section 3, I describe data retrieval and processing. Sections 4 through 7 present the results of the study, including the pairwise correspondences between the verbs under analysis, the similarity of various sight verbs to basic verbs of the class, the distribution of correspondences to non-basic verbs, and the semantic groupings of the verbs. Section 8 summarizes the main findings of the study.

2. Semantic distinctions and polysemy patterns of sight verbs

This study aims to establish the semantic similarity between sight verbs and their groupings bottom-up, as the result of the analysis of the data. Still, it may be instructive to bear in mind the semantic distinctions that, based on the literature, can be expected to play a role in the structuring of the sight domain. Knowing the polysemy patterns of sight verbs can help anticipate the possible sources of variation in how the verbs are used. Also, since it will be impossible to analyze the semantic relationships between the verbs under analysis in detail, this section provides a look beyond the frequencies of correspondence into the actual use of verbs in the context.

The semantic distinctions structuring the domain of sight can either be domain-specific or belong to more general semantic and grammatical distinctions, and among the latter, aspect is the most prominent. Differences in aspectual behaviour can naturally affect the distribution of verbs across contexts and thus the degree of correspondence between verbs across languages, see (Wälchli 2016). For instance, there are considerable differences in the aspectual behaviour of Bulgarian and Russian basic experience verbs of sight and hearing: B *виждам — видя* and *чувам — чуя* and R *видеть — увидеть* and *слышать — услышать*. In particular, in general-factual contexts, such as (3), Bulgarian uses the perfective verbs, whereas Russian strongly prefers the imperfective, see (Ovsjannikova, Kukova 2022; 2023) for a Bulgarian-Russian parallel corpus study and (Wälchli 2016: 86).

(3)   *Meysan, siguren li si, che si **vidyal** podvodnitsa?*
        *Jesteś pewien, że **widziałeś** łódź podwodną ?*
        *Mejson, ty točno **videl** podlodku?*

'Mason, are you sure you saw a sub out there?'

This difference might be a manifestation of a global difference between the two languages, i.e., of a general propensity of Russian imperfectives to be used in general-factual contexts, as discussed by Knjazev (2007, 379–384). Still, in these pairs this propensity is much more pronounced than in verbs of other groups, e.g., in activity verbs of sight and hearing.

Generally, this study does not delve into the differences in the aspectual behaviour of individual verbs. However considerable they may be, the study shows that, with very few exceptions, aspect remains one of the major parameters in structuring the verbs under analysis. Specifically, a higher degree of correspondence is consistently observed between verbs of the same aspect, and further semantic distinctions are mostly drawn within the imperfectives and the perfectives. This generalization may not hold as strongly for Slavic as a whole, because the selected languages do not lie on the opposite poles in terms of aspect usage. In particular, in the east-west split suggested by Dickey (2000) both Bulgarian and Russian belong to the east group and Polish is considered intermediate between the two, see also von Waldenfels (2012) on the use of aspect in imperative.

One of the distinction specific for the domain of perception and central to it is that between experience and activity verbs. It is typically discussed with respect to basic verbs of perception, but even for such verbs this distinction turns out to be considerably blurred. One of the issues is the semantic content of this distinction, which is often taken to be self-explanatory in works on perception, but in reality, does not necessarily boil down to the presence vs. absence of control, see the discussion in Wälchli (2016: 71-72). The variation observed in the distribution of experience and activity verbs in parallel contexts suggests that this distinction is not realized uniformly across languages and that there are types of perception events prone to more variation, such as ambulatory vision ('go/come and see') discussed by Wälchli (2016: 72, 79). The present study reconsiders this distinction throwing a wide range of non-basic sight verbs into analysis and confirms its relevance for the structuring of the sight domain.

Basic perception verbs serve as the default means to describe a certain type of perception, and as such they are expected to have a higher token frequency as compared to other verbs denoting the same sense and to be semantically general, in the sense that they describe the respective perception type without specifying, e.g., its manner or duration (San Roque et al. 2015: 40). Some of non-basic sight verbs are as semantically general as basic verbs as far as the type of perception they denote is concerned but more or less restricted in terms of register or grammatical features. An example is the Russian verb *gljadet'*, which is fully synonymous to the basic verb *smotret'* 'look' and has a very similar distribution of grammatical forms, except that *gljadet'* is much more frequently used as a converb (*gljadja*) and is mostly attested in fiction. Semantic generality can also be regarded as a matter of degree. Diachronically, some verbs with more specific meanings start to be used in a wider range of contexts,

whereby their specific semantic features can gradually fade. For instance, the Polish verb *spojrzeć* is usually defined as denoting a brief glance. However, presently, this verb is by far the most frequent perfective activity verb of sight, which suggests that it should be regarded as one of the basic verbs and that the meaning component of brief perception might be no more as highlighted. In sections 5 and 6 below, the issue of semantic generality is addressed from a parallel corpus perspective.

Grammatically, additional semantic specification of perception in Slavic is mostly related to various Aktionsarten, expressed by prefixes (Paducheva 2004: 198; Nesset 2010). The semantic components these verbs bear include brief and long perception, intensive and complete perception. Wälchli (2016: 56, 99) discusses obscured perception verbs, such as R *razgljadet'*, which denote seeing something with effort, despite bad conditions. Prefixed verbs also typically express various spatial configurations accompanying sight, such as looking around, over, or from behind an obstacle. The same prefixed verb can often highlight different aspects of the perception event at the same time. For example, the Bulgarian verb *ogleda* can convey both the meaning of looking around (4) and fully perceiving an object (5), and often it is not easy to distinguish between the two.

(4)   *Toy nervno **ogleda** pustite halmove.*
        ***Rozejrzał się** nerwowo po okolicznych pustych wzgórzach.*
        *Leonardo ispuganno **ogljadelsja** po storonam.*
        'He looked nervously around the deserted hills.'

(5)   *No posle **ogledah** lodkata im.*
        *Ale **przyjrzałem się** ich łodzi.*
        *Potom ja horošen'ko **prismotrelsja** k ih lodke.*
        'Then I got a good look at their boat.'

Due to the high number of prefixed verbs and the differences in their usage patterns across the three languages, establishing a clear typology of these verbs is challenging. For this reason, investigating the groupings of these verbs that emerge from the data can be particularly important and insightful.

Polysemy patterns and meaning extensions are usually discussed in relation to basic verbs of perception (San Roque et al. 2018; Georgakopoulos et al. 2021), but they are also relevant for the present study, because the meanings expressed by basic verbs of sight in one language can be taken up by non-basic verbs in another language. In a parallel corpus, this results in a higher frequency of correspondences between the basic verb in one of context types and the non-basic verb, thus contributing to semantic similarity between them. For instance, one of the polysemy patterns attested for perception verbs is when they are employed to express the meaning of co-identification, which can be paraphrased by English expressions *consider to be* or *regard as* (San Roque et al. 2018: 380). A

recurrent correspondence observed for this meaning in the languages under analysis is that Bulgarian uses the basic experience verbs of sight, Polish uses the basic activity verb of sight, and Russian uses the non-basic verb *rassmatrivat'*, originally expressing the idea of attentive inspection, as in (6).

(6) *Kato pri vsyako izmenenie na vremeto, nie okurazhavame horata da **vidyat** Umnia Dazhd dalgosrochno podobryavane na sistemata za kulturata.*
*Jak z każdą modyfikacją pogody, chcemy by ludzie **patrzyli** na Inteligentny Deszcz jako na rozwiązanie długoterminowe, a nie na szybką naprawę.*
*Kak s ljuboj sistemoj modifikacii pogody, my prosim ljudej **rassmatrivat'** umnyj dožd' kak dolgosročnuju sistemu obogaščenija urožaja.*
'As with any weather modification system, we encourage people to view smart rain as more of a long-term crop enhancement system.'

Meaning extensions of sight verbs to other domains, such as cognition ('understand', 'deduce') and attention ('examine', 'check'), see (San Roque et al. 2018: 380), can manifest in correspondences between these verbs and verbs of other semantic classes in parallel contexts. For some verbs, more than half of the occurrences in parallel texts falls on the correspondences with verbs outside of the domain of sight, e.g., the Russian verb *rassmatrivat'* much less frequently corresponds to Bulgarian and Polish verbs of sight than to verbs meaning 'examine', 'consider' and the like. In section 3, I show to what extent the occurrences of the sight verbs in each of the languages overlap with the sight verbs in other languages.

Setting aside the recognized meaning extensions, the situations describing sight proper in reality fall into a variety of subtypes, such as watching a movie, looking into a mirror and looking into someone's eyes, or more abstract situations of looking into someone's soul or heart or looking into the future. In a particular language, some of these minor subtypes of sight can call for a particular verb, different from the basic one, which also affects the degree of correspondence between verbs in parallel corpus data. Let us take a situation of looking out of the window as an example, see (7).

(7) *Tya se razhozhdashe iz kabinata, nadnichashe vav vsichki agli, **pogledna** prez prozoretsa*
*Chodziła po kabinie, zaglądała we wszystkie kąty, **wyjrzała** przez okno*
*Ona hodila po kabine, zagljadyvala vo vse ugly, **posmotrela** v okno*
'She walked to and fro, looked into all corners, looked out of the window.' (author's translation)

Table 1 shows the distribution of the major sight verbs used with the phrase 'out of the window' in the three languages[2]. The upper row shows Polish verbs, and the columns show the frequencies of their correspondence to Russian and Bulgarian verbs (the correspondences between the latter two languages are not shown). The verbs are ordered by frequency; perfective verbs are underlined.

Table 1. Verb correspondences for the situation of looking out of the window

	wyjrzeć	*patrzyć*	*wyglądać*	*spojrzeć*	*gapić się*	*popatrzyć*	*zaglądać*	**Sum**
pogledna	39	2		6		2		**49**
gledam		15	8	1	3		2	**29**
zagledam se	1	1	1	1		1		**5**
poglezhdam	2		2					**4**
zjapam		1			1			**2**
nadnicham			1	1				**2**
vziram se					1			**1**
Sum	**42**	**26**	**12**	**9**	**5**	**3**	**2**	**99**
vygljanut'	26	2		5		1		**34**
smotret'	3	19	8		1		1	**32**
posmotret'	12	3	1	4			1	**21**
gljadet'		1			2	2		**5**
ustavit'sja		1	2					**3**
pjalit'sja					2			**2**
vygljadyvat'	1		1					**2**

The majority of uses in Table 1 are distributed between basic activity verbs of sight and verbs which convey an idea of looking from behind an obstacle. The three languages behave differently as regards the distribution between these types. Polish preferably uses the verbs specifying the spatial configuration, especially for the perfective aspect (*wyjrzeć*). In Bulgarian, semantically general verbs are predominantly used in this context, whereas the other verbs are much less frequent. Finally, the distribution in Russian is more similar to that in Polish: among perfective verbs, the semantically specific verb *vygljanut'* is preferred, but the imperfective *vygljadyvat'* is less frequent than its Polish counterpart *wyglądać* (as compared to the respective perfectives; two-tailed Fisher test, $p \approx 0.04$).

Table 1 also shows that the only feature where the correspondences between verbs are relatively consistent is aspect: imperfectives and perfectives typically correspond to verbs of the respective aspect

---

[2] The query included the phrases B *prez prozoretsa*, P *przez okno*, and R *в окно*. Then I manually selected the contexts describing the situation of looking out of the window from inside. Contexts with experience verbs were excluded. In Table 1, the occurrences of the lemmas *patrzeć* and *patrzyć* are considered as the same verb.

in another language. Otherwise even within this semantically restricted situation type there is considerable variation in the choice of verbs. Even more complex and multifaceted picture emerges when the whole range of uses is considered together for several dozens of verbs in each of the languages.

3. Data retrieval and processing

The major source of data for this study is the parallel corpus InterCorp (Rosen et al. 2022). The searches were conducted using bilingual parallel subcorpora, i.e., each time searching in all the texts available only in two of the three languages. Although a different subcorpus size and a slightly different set of text types is available for the three pairs of languages, the majority of texts in all the three subcorpora belong to subtitles, as shown in Table 2.

Table 2. Size and composition of the three bilingual subcorpora in InterCorp (as of September, 2024)

Language pair	Bul-Pol	Bul-Rus	Pol-Rus
Size	223.5 mln tokens	115 mln tokens	136 mln tokens
Composition	Subtitles 85% Legal texts 8% Discussions' transcripts 5% Fiction 3%	Subtitles 96% Fiction 4%	Subtitles 92% Fiction 7% Other 1%

As pointed out by Levshina (2016: 516), subtitles represent spoken discourse and spontaneous conversations, which are only marginally present in fiction and other text types. For the present study, this means that the data may contain more discourse uses of sight verbs, e.g., imperative forms employed to direct attention or manage interaction (San Roque et al. 2018). In other respects, parallel subtitles have been shown to be a reliable source of data for language comparison, despite their translational nature and the specific conditions of their creation and use (Levshina 2017).

Data retrieval and processing included the compilation of the lists of verbs, the retrieval of frequencies for all the pairs of verbs, and the correction of the data for false correspondences.

First, a list of verbs was compiled for each of the languages under analysis. The lists include experiencer-subject verbs describing visual perception[3]. The lists were mostly compiled bottom-up, based on the verbs attested in the searches. However, for the list of Russian verbs, I also used a frequency list created on the basis of the semantic annotation of the Russian National Corpus and described in (Ovsjannikova 2019). The lists were intended to be as exhaustive as possible. However, even apart from the fact that in language use, new verbs can emerge every second making it impossible to create a truly exhaustive list of any semantic class, a frequency threshold had to be introduced. Only

---

[3] The lists do not include the verbs denoting lack of perception, such as P *przeoczyć* 'overlook' and the verbs with the nouns denoting eyes as the direct object, such as B *vtrencha* 'stare'.

the verbs having more than 100 occurrences in all of the bilingual corpora were included. As a result, the lists included 34 Bulgarian, 38 Polish, and 42 Russian verbs.

At the next stage, the frequencies of correspondence between all pairs of verbs for the three pairs of languages were found, using lemma search. Unfortunately, the frequencies yielded by the searches cannot be taken at face value, and the number of false correspondences may be very different for different pairs. False correspondences are typically observed when two verbs are used in a sequence of perception events, where one verb describes direction of attention and the other the resulting perception event, as in (8). Another frequent configuration leading to false correspondences features one experiencer perceiving the perception event by another experiencer, as in (9).

(8)   *Frodo **se ogleda** nazad i **zarna** otblyasaka na byala pyana sred sivite darvesni stabla.*
        *Frodo **obejrzał się** i **dostrzegł** blask białej piany między szarymi pniami drzew.*
        *Frodo **ogljanulsja** i **uvidel** sredi drevesnyh stvolov beluju penu vodopada.*
        'Frodo looked back and caught a gleam of white foam among the grey tree-stems.'

(9)   *I dokato nikoy ne **gledashe**, **vidyah** kak tya **pogledna** kam teb.*
        *Gdy nikt nie **patrzył**, **widział em**, jak na ciebie **zerka**.*
        *I kogda nikto ne **videl**, ja **zametil**, kak ona **smotrit** na tebja.*
        'And when no one else was looking, I saw the way she glanced at you.'

Such examples depict typical scenarios of visual perception and as such can provide a richer understanding of the semantics of sight verbs and probably of the processes of visual perception in general, which remains beyond the scope of this study.

To correct the data for false correspondences, the following procedure was implemented. If the number of hits for a pair of verbs was less than 100 the examples were manually analysed and the exact number of correct matches was counted for the pair. If the number of hits was greater than 100 the search results were shuffled and the proportion of correct hits was counted for the first 100 occurrences. Then the number of hits shown in the corpus was multiplied by that proportion.

Additional processing of corpus results is also necessary in case of verb stems which can be used both with and without a reflexive marker, such as B *zaglezhdam* (*se*). In this respect, one of the grammatical differences of Bulgarian and Polish from Russian is that in the latter two languages, the reflexive marker is a clitic written separately from the verb, and in Russian this marker is a suffix which is written together with the verb. Therefore, for Russian, the same stem with and without the reflexive marker is treated as two different lemmas in the corpus. For the other two languages, the implemented approach was to estimate the number of reflexive and non-reflexive uses based on the first 100 examples or count their frequencies among all the examples, if less than 100.

9

Reflexive markers in all the three languages have a wide range of functions, including reflexive proper (in a broader or narrower sense), reciprocal and passive, see (Knjazev 2007; other references?). Reciprocal and passive verbs were not included in the list for Russian and such uses were excluded from the counts of correspondences for Bulgarian and Polish. Note that this did not affect passive participles, as participles are retrieved in the searches of the verbs as their morphological forms.

The general criterion for identifying a pair of parallel contexts as a correspondence was the identity of construction type, i.e., subject-experiencer, and the semantic identity of participants. The latter is especially relevant for cases where the perceived object is expressed in different ways in the two languages, as in (4)-(5) above.

Having retrieved and corrected the frequencies of correspondence for each pair of verbs, we can examine to what extent the verbs of one language overlap with each of the verbs of the other language. For example, we can sum up the frequencies of correspondence of all Polish verbs in the list to the Bulgarian verb *nablyudavam* 'observe' and calculate the proportion of the uses of this verb in the Bulgarian-Polish subcorpus "covered" by the Polish verbs. By calculating this proportion for all the pairs across the three languages, we can also assess whether the difference in the number of verbs chosen for the three languages affects the degree of this coverage.

Figure 1. Proportions of uses of the verbs of each language covered by the sight verbs of the other two languages
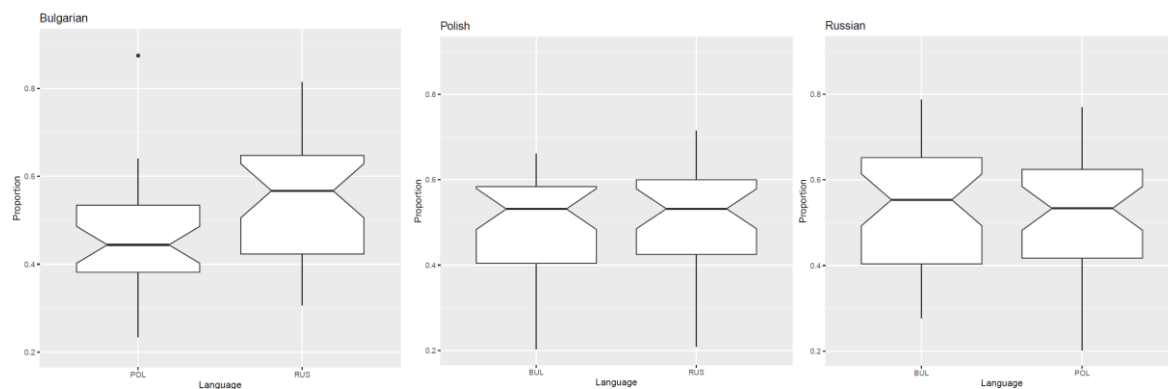
Bulgarian is the only case where there is a significant difference between the two languages in terms of the median proportion of covered uses (as shown by the notches on the boxes). However, this is unlikely to be explained by the fact that the list of verbs includes more verbs for Russian than for Bulgarian, because in the other two cases the medians of the distributions are very close. It is especially noteworthy that the 34 Bulgarian verbs cover the 42 Russian verbs slightly better than the 38 Polish verbs. This suggests that Russian and Bulgarian may be more similar to each other in terms of the range of uses of sight verbs than to Polish. More importantly for this study, there is no consistent evidence that the

difference in the number of verbs included in the lists for the three languages affects the proportion of uses covered by them in the other languages.

4. Mutual pairwise correspondences: Dice-coefficient

As the first take at the semantic similarity between the verbs, for each pair of languages, we can find the pairs of verbs which show the highest degree of correspondence to each other. This degree was measured using Dice-coefficient, which is used in collocation analysis to express the degree of attraction between collocates, i.e., words that frequently cooccur together in text (Evert 2009). The Dice coefficient for a pair of co-occurring words is calculated by multiplying the frequency of their co-occurrence by two and dividing it by the sum of their cooccurrences with all other words (which equals their total overall frequencies).

In the case of correspondences in parallel texts, the denominator included the sums of correspondences to all the verbs of the other language.

Tables 1A, 2A, and 3A in the Appendix show the pairs of verbs with the highest degree of correspondence for the three pairs of languages as estimated using Dice-coefficient. The resulting pairs cannot be viewed as the best translational equivalents, at least for less frequent verbs, as their potential better equivalents can be absent from the lists.

In general, the verbs in pairs systematically match with respect to aspect, with only several exceptions. Mismatches in reflexivity are more numerous, which is understandable, since the choice of aspect is much more directly triggered by the context and reflexivity has more to do with the verb's derivational history. More on this? Examples?

Some verbs occur in the tables more than once. This means that a verb in one language happen to have the highest value of Dice-coefficient for more than one verb in another language. This is inevitable, because in all the three pairs the list of verbs for one of the languages has fewer verbs than that for the other, but such cases are also observed for the languages with the larger lists (Russian in both pairs, Polish as compared to Bulgarian)[4]. This suggests that these recurring verbs have a higher degree of semantic generality, since their range of contexts is shared between at least two verbs in the other language. For instance, one of the Russian verbs that recurs in both tables is *razgljadet'* 'make out'. This verb shows the highest value of the Dice-coefficient for P *przyjrzeć się* 'take a close look' and *dojrzeć* 'glimpse', see Table 3A. Among the pairs of Bulgarian and Polish verbs, the verb *wpatrzywać się* is the best correspondence for the verbs *vziram se*, *zagledam*, and *vtrencha se* (all of them can be translated by the English verb *stare*), see Table 1A. It is likely that the Polish verb conveys the meaning of a long attentive gaze in a more general way than the Bulgarian verbs.

---

[4] Technically, as the first step, the pairs with the highest Dice-coefficient were found for the language which had a larger number of verbs, e.g., for each Russian verb a Bulgarian counterpart was found. Then I checked which verbs of the second language, in this example Bulgarian, are absent from the already found pairs and found the Russian pairs for them.

Basic sight verbs do not generally recur in the tables, although, as discussed in section 5 below, many non-basic verbs of other languages frequently correspond to them. However, due to their very high frequency, which goes to the denominator in the formula of the Dice-coefficient, pairs of basic and non-basic verbs show comparatively low degree of correspondence. Thus, the recurrence of verbs in the tables mentioned above helps identify verbs which are general for specific subtypes of sight rather than for the whole domain.

In all the three tables, the basic verbs of sight, especially experience verbs, as well as the verbs 'notice' and 'observe' are among the highest-ranking pairs. They are also among the most frequent verbs in the three languages. This raises the question of whether there is a correlation between the verbs' frequency and their degree of correspondence, given that they are already established as a semantically best-matching pair.

Theoretically, two alternative hypotheses can be put forward concerning the relation between frequency and mutual correspondence between verbs in a language. One possibility is that more frequent verbs, which are also likely to be more semantically general, are more similar to each other across languages, whereas less frequent members of the class exhibit more irregularity and carve the semantic space in more idiosyncratic ways. Alternatively, the core members of the class may be more prone to semantic extensions, which need not be the same across languages, while the more specific meanings of less frequent verbs might more neatly correspond to each other.

To test for the presence and the sign of the correlation, one has to take into account that Dice-coefficient is not independent of the verbs' frequency. First, verb pairs containing one very frequent verb and one verb of low frequency cannot by definition have a high Dice-coefficient, because the frequency of their correspondence cannot in any case be higher than the frequency of the less frequent verb. Second, and even more importantly, more frequent verbs can be expected to share a larger part of their contexts by chance. For these reasons, the correlation between the Dice-coefficients and the verbs' frequency cannot be tested directly. Instead, for each verb pair, one can compare the observed Dice-coefficient with a distribution of Dice-coefficients resulting from a large number of randomly generated correspondence frequencies5. Then the observed Dice-coefficient is expressed as a standardized z-scores, based on its deviation from the mean Dice-coefficient based on randomly generated frequencies. The standardized z-scores can be tested for correlation with the frequency of the verbs in a given pair. The natural logarithm of the frequency of the less frequent verb was used as a measure of the frequency for a pair of verbs. For all the pairs of verbs, a significant positive correlation is observed between the z-scores of the observed Dice-coefficients for the pairs of the best matching verbs and the frequency of the less frequent verb of the pair ($r(48) = 0.65$, $p < 0.0001$ for Bulgarian and Polish; $r(49) = 0.68$, $p <$

---

[5] Technically, 1000 tables with random distributions of correspondence frequencies were generated, keeping constant the row and column sums of the original correspondence table.

0.0001 for Bulgarian and Russian; r(53) = 0.59, $p < 0.0001$ for Polish and Russian).[6] These correlations are shown in Figure 2.

Figure 2. Correlations between the verbs' frequency and the z-score of the pair's Dice-coefficient
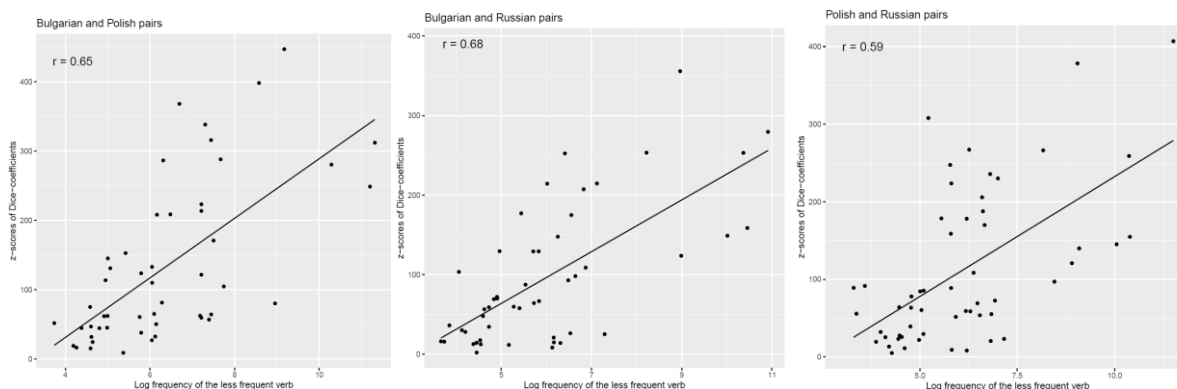
These correlations favor the hypothesis of the positive correlation between the verbs' frequency and the strength of the mutual correspondence. This correlation may suggest that the mutual correspondence between more frequent verbs is not so much affected by differences in the minor patterns resulting from language-specific semantic extensions, whereas for less frequent verbs, even minor discrepancies in usage patterns can result in a considerably lower degree of mutual correspondence. Consider again the correspondences between the Polish verb *wyjrzeć* and the Russian verb *vygljanut'*, discussed in section 2. Although these two verbs seem to match very closely both derivationally and semantically, their mutual correspondence, at least in the context of looking out of a window, is affected by the fact that Russian more often uses basic verbs of sight for this type of context or, put differently, is less prone than Polish to specify the spatial configuration in this case. In the next section, I look at the extent to which the verbs under analysis correspond to basic sight verbs and how this correspondence can be interpreted in terms of semantic generality.

5. Similarity to basic sight verbs

As discussed in the previous section, the Dice-coefficient takes into account the overall frequency of each member of the pair. This helps detect pairs of verbs which have the highest degree of mutual correspondence. However, for some lower-frequency verbs, the correspondences with one of the basic verbs of sight are also frequent, in some cases even more frequent than with the verb closest to it in terms of Dice-coefficient. Example?

The high frequency of correspondence of more semantically specific verbs to basic verbs of sight is not surprising, since basic verbs of sight denote sight in its most general form and are stylistically

---

[6] Significant correlations are also observed between the standardized Dice-coefficients and: a) the mean logarithm of frequency and b) logarithm of mean frequency for verb pairs with comparable frequencies (where the frequency of one verb is more than 0.5 and less than 1.5 of the frequency of the other).

neutral. It can also be expected that different verbs of sight will be more similar either to experience or to activity sight verbs. In this section, I explore this question by using basic verbs of sight as two poles of the sight domain and as ideal representations of the experience vs. activity distinction and by looking at the degree to which various sight verbs correspond to basic verbs.

For instance, for each Bulgarian verb, I calculated the proportions of uses where it corresponds to the basic experience sight verbs in Polish and in Russian and the mean of these proportions, as well as the proportion of uses where it corresponds to the basic activity sight verbs in these languages and the mean of the two[7]. These mean proportions for each of the three languages are visualized in a two-dimensional graph, see Figure 3. The x-axes of the graphs correspond to the mean proportions of correspondences with basic experience verbs and the y-axes, with activity verbs. The verbs that are close to the origin, i.e., are located in the lower left corner of the graph, rarely correspond to either basic experience or activity verbs.

Figure 3. Proportions of correspondences to basic verbs

---
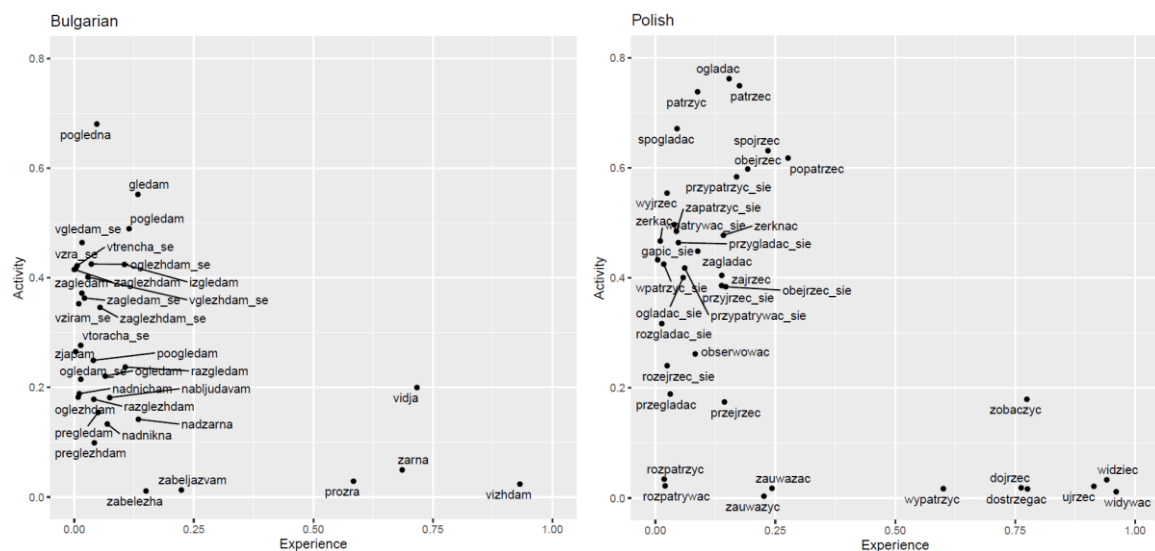
[7] This approach may have its drawbacks, because basic verbs of sight in each of the languages may have idiosyncratic patterns of use and different semantic ranges. This problem is somewhat alleviated by using two languages as the basis of comparison. A more language-neutral estimates can arguably be obtained by using a larger number of languages for comparison, as then idiosyncratic properties of individual verbs will be more levelled out.
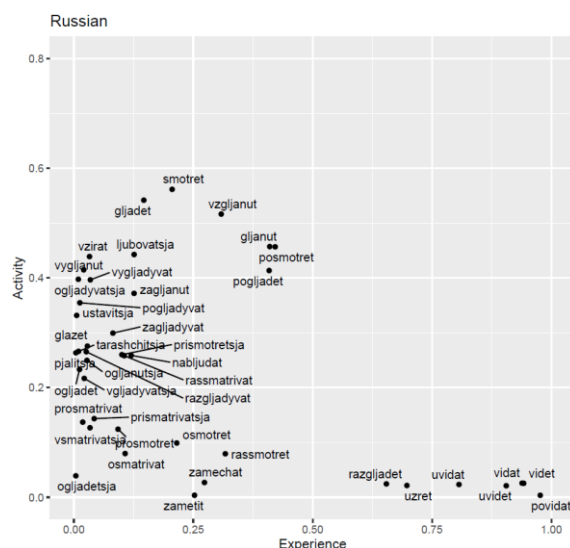
The general pattern shown by graphs in Figure 3 is that the verbs are mostly located along the axes and the centres of the graphs remain empty. This means that the verbs gravitate toward the experience pole or the activity pole or to neither of them, but rarely share equally large parts of their contexts both with experience and activity verbs. Thus, the distinction between experience and activity may indeed be viewed as structuring the domain of sight, including its non-basic members, with no intermediate zone between the two poles.

The most notable exceptions from this general pattern are the basic experience verbs *vidja* in Bulgarian and *zobaczyć* in Polish and the basic activity verb *posmotret'* in Russian with a few non-basic verbs located close to it. The correspondences between these verbs are mainly observed in the contexts shown in (10) and (11).

(10) *no **vizhte** kakvi sa techeniyata po shelfa sega.*
      *ale **spójrz** na prądy wzdłuż szelfu teraz.*
      *I **vzgljanite** teper' na tečenie vdol' berega.*
      'But look at the currents along the shelf now.'

(11) *Shte **vidya** ako mozhem da prosledim tezi hora.*
      Zobaczę, czy możemy wyśledzić tych ludzi.
      *Posmotrju**, smožem li my otsledit' ètih ljudej.***
      'I'll see if we can't track these people down.'

In imperative contexts (10), Bulgarian systematically uses the experience verb *vidya* 'see' in contrast to the other two languages, and in first person future contexts (11) Russian is opposed to the other to languages as it much more frequently uses the activity verb *posmotret'* 'look'. Uses such as these are found in dialogical interaction and may be viewed as semantic extensions of the basic sight meaning. As such extensions are arguably more typical for basic than non-basic verbs, the basic experience and

15

activity verbs paradoxically show more similarity to each other than non-basic verbs lying close to one of these poles.

Figure 3 also shows that many more verbs tend toward the activity pole than the experience pole. This is reflected in the maximal proportions on the two axes. It is much closer to 1 in case of experience verbs, because they mostly correspond to themselves, whereas the basic activity verbs have many correspondences to non-basic verbs. This picture suggests that basic activity verbs are less diachronically stable and more to prone to renewal than basic experience verbs. This is also supported by the cognate relations between the basic experience verbs in the three languages and the etymological diversity of the basic activity verbs. As shown in Figure 3, some of the non-basic verbs, especially in Polish and Russian, are very close to the basic activity verbs, as *oglądać* in Polish and *vzgljanut'* in Russian, denoting full perception and brief glance, respectively. In the next section, I show that these verbs are close to the basic activity verbs also in terms of frequency and the distribution of their correspondences to other verbs.

At the experience pole, there are also verbs that group close together with the basic verbs, such as the verbs *vidat'*, *uvidat'*, and *povidat'* in Russian. Although these verbs can be said to express perception as generally as basic verbs do, they have low frequency and are aspectually and/or stylistically restricted as compared to the basic verbs. Thus, they cannot be regarded as rivals to basic verbs, as the abovementioned verbs at the activity pole.

6. Distribution of correspondences to non-basic verbs: Shannon entropy index

One of the manifestations of semantic generality and contextual neutrality of basic verbs is that they can correspond to a wide range of non-basic verbs, irrespective of their specific semantic features. In the previous section, this was shown by plotting for the verbs of each of the languages the proportions of correspondences to basic verbs of the other two languages.

Non-basic verbs can also differ with respect to the distribution of correspondences to the verbs in the other languages. Some verbs mostly correspond to one or two verbs in another language, whereas the uses of some verbs are more evenly distributed among the verbs of another language. For instance, the Bulgarian verb *zyapam* and the Polish verb *gapić się* 'stare', while showing a high degree of mutual correspondence, have different distributions of correspondences. This is shown in Table 3, which contains the seven verbs (Polish and Bulgarian, respectively) with the highest frequencies of correspondence to the verbs *zyapam* and *gapić się*. The table shows raw correspondence frequencies and their proportions.

Table 3. Polish verbs corresponding to *zyapam* and Bulgarian verbs corresponding to *gapić się*

Polish verb	Raw frequency	Proportion	Bulgarian verb	Raw frequency	Proportion
gapić się	1231	0.72	*gledam*	1451.12	0.45
patrzeć	225.79	0.13	*zyapam*	1231	0.39
patrzyć	70	0.04	*vziram se*	229	0.07
wpatrywać się	42	0.02	*nablyudavam*	57	0.02
oglądać	37	0.02	*vtoracha se*	54	0.02
przyglądać się	33	0.02	*pogledna*	26	0.01
obserwować	15	0.01	*oglezhdam*	26	0.01

The verb *zyapam* corresponds to the verb *gapić się* in more than two thirds of its uses, whereas the verb *gapić się* shares the majority of its uses between the verbs *gledam* and *zyapam*. This suggests that the Polish verb is more stylistically neutral than its Bulgarian counterpart.

The extent to which the distribution is concentrated within a small number of values or more evenly spread among a larger number of values can be measured with the Shannon entropy index, see, e.g., Stoll et al. (2017) for its application in linguistics. The higher is the entropy of the distribution, the more evenly spread it is among the values. For example, for the verbs *zyapam* and *gapić się*, the Shannon entropy index based on the distribution of all the correspondences is 1.99 and 2.18, respectively.

The proportions of correspondences to the basic verbs were analysed in the previous section and these proportions can considerably affect the entropy index. For these reasons, only the correspondences to non-basic verbs are taken into account in this section. For instance, for each Bulgarian verb (including the basic ones) the Shannon entropy index was calculated based on the distribution of uses for all the non-basic Polish and Russian verbs (using the R package *entropy* (Hauser, Strimmer 2009)). In Figure 4, the values of Shannon entropy index for Bulgarian, Polish, and Russian verbs are plotted against the logarithm of the verbs' frequencies (the mean sum of correspondences to all the verbs in the two languages).

Figure 4. The entropy of the correspondences to non-basic verbs and verbs' frequency
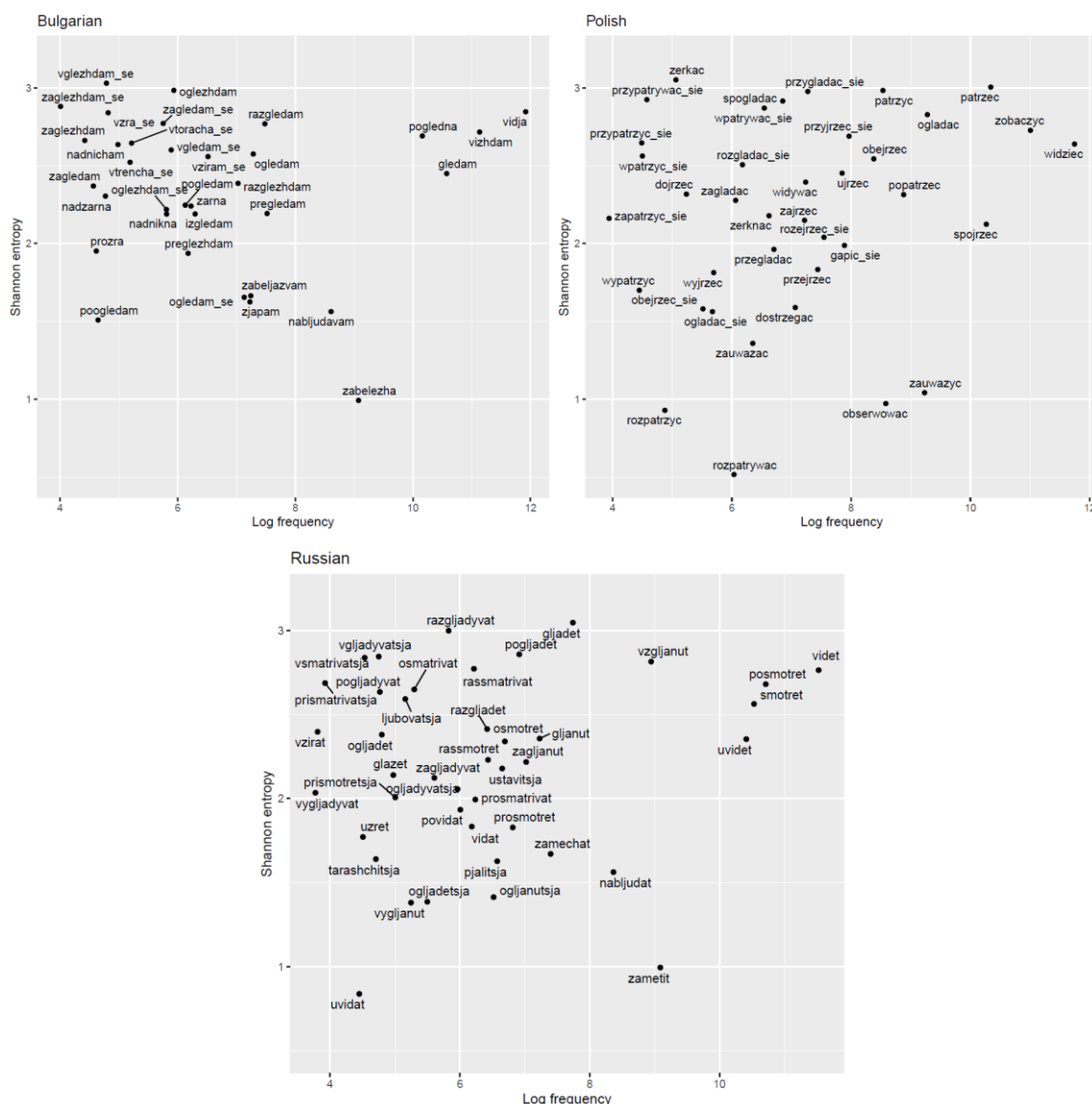
In Figure 4, the basic sight verbs are found in the top right part of the graphs: they have both high frequency and high values of entropy, which means that their correspondences to non-basic verbs are considerably spread. Verbs close to the basic verbs in terms of frequency and entropy may be regarded as their (potential) rivals. Such verbs are especially numerous in Polish, e.g., *oglądać* and *popatrzeć*, testifying to a less clear-cut distinction between basic and non-basic verbs in this language. In Russian, the verb *vzgljanut'* 'take a quick look' is close to the area of the basic verbs. All these Polish and Russian verbs also show a high proportion of correspondences to basic activity verbs, as shown in Figure 3 in section 5. In Bulgarian, there are no verbs as close to the basic verbs in terms of frequency and entropy. For all the three languages, a significant positive correlation is observed between entropy of correspondences to non-basic verbs and the proportion of correspondences to activity verbs ($r(32) = 0.43$, $p \approx 0.01$ for Bulgarian; $r(36) = 0.57$, $p < 0.001$ for Polish; $r(40) = 0.38$, $p \approx 0.01$ for Russian). No significant correlation was found between frequency and entropy.

18

The top left parts are densely populated in all three graphs. Many of the verbs found here denote focused or thorough visual perception and their correspondences are considerably spread between several verbs of this group, suggesting less clear semantic distinctions between them. As an illustration, Table 4 shows the correspondence frequencies for several Polish and Russian verbs from this group and their entropy values based on the correspondences to non-basic verbs (underlined are the entropy values for imperfectives).

Table 4. Correspondence frequencies and entropy values for several Polish and Russian verbs of focused and thorough perception

Verb	*przyjrzeć się*	*przyglądać się*	*wpatrywać się*	*wpatrzyć się*	Entropy of Russian verbs
rassmotret'	240.57	7	0	0	2.23
razgljadet'	131.52	4	0	0	2.41
rassmatrivat'	26	90	12	2	<u>2.77</u>
razgljadyvat'	20	74	20	1	<u>3.00</u>
prismotret'sja	99.99	6	1	0	2.01
ustavit'sja	8	16	59	16	2.18
vgljadyvat'sja	5	12	35	6	<u>2.85</u>
vsmatrivat'sja	1	13	31	1	<u>2.84</u>
prismatrivat'sja	8	17	0	0	<u>2.69</u>
Entropy of Polish verbs	2.70	<u>2.98</u>	<u>2.87</u>	2.56	

Although the distributions shown in Table 4, as viewed per rows and per columns, are far from uniform, there are typically not clearly skewed in favour of just one or two verbs, either. One of the reasons for this is the existence of synonymous Russian verbs with the roots *gljad-* and *smotr-* (similar pairs can also be found in Polish for the roots *patr-* and *gląd-*). However, other verbs also contribute to the fuzziness of the correspondences in this group. In particular, the Polish verbs with the prefix *przy-* most frequently correspond to the Russian verbs with the prefix *raz-*, but they also share some contexts with the derivationally cognate verbs with the prefix *pri-*. For the verb *wpatrywać się*, the highest frequency of correspondence is observed with the verb *ustavit'sja*, but there are also many contexts it shares with other verbs in Table 4.

Naturally, there is no clear-cut distinction between the groups of verbs outlined above based on the graphs in Figure 3. Verbs with higher entropy values and higher frequency are expected to have a wider range of contexts and be less semantically specific, and thus be closer to basic verbs in terms of semantic generality and contextual neutrality.

The verbs with lower entropy values found in the bottom parts of the graphs vary in frequency but they all more or less closely correspond to one verb in each of the remaining languages, as in cases of the verbs *nablyudavam*, *obserwować* and *nabljudat'* or *ogledam se*, *obejrzeć się*, *ogljadet'sja* and

ogljanut'sja. These verbs also mostly have low proportions of correspondences to basic verbs, as shown in Figure 3. Thus, verbs from the lower part of Figure 4 are more semantically specific than those in the top parts of the graphs.

7. Semantic groupings in the sight domain

As shown in the previous sections, apart from the pairs of verbs with the highest degree of mutual similarity, measured here using Dice-coefficient, there are more or less strong relations of mutual correspondence on other levels, whereby the majority of verbs under analysis show some degree of correspondence to a number of verbs in another language. Sight verbs within one language can also be more or less similar to one another. Visualization and analysis of these complex relations is the central topic of this section.

To explore the groupings of verbs both within and across languages, it is desirable to be able to measure similarity between verbs within one language and the degree of correspondence between verbs in different languages in the same or in a similar way. In this case, the following approach was implemented. Similarity between verbs of the same language was measured comparing the distributions of their correspondences to the verbs in the other two languages. For example, for B *nadnikna* and *nadzarna* 'look over', I compared whether they have similar distributions of correspondences to all the Polish and Russian verbs under analysis. The degree of correspondence between verbs in two different languages was compared based on the distributions of their correspondences to the verbs in the remaining language. For instance, B *nadnikna* 'look over' and R *zagljanut* 'look in' were compared in terms of the similarity of their correspondences to the Polish verbs. Similarity between the distributions under comparison was measured using the cosine similarity, which is widely used to compare documents on the basis of words they contain (Singhal 2001). Then cosine similarity was turned into distances between verbs and a multidimensional scaling (MDS) algorithm was used to visualize these distances (with the help of the R package *smacof* (Mair et al. 2022)). MDS algorithm aims to visualize distances in a low number of dimensions with the minimal distortion of original distances. The extent of this distortion is called "stress", and its value should be as low as possible to ensure a reliable visualization of objects. In the case of the sight verbs under analysis, the 3-dimensional visualization had to be chosen, with the stress value of 0.19, which is high but still acceptable (Levshina 2015: 341). Figure 5 shows the MDS graph visualizing all the 114 verbs under analysis. The imperfective verbs are plotted by white points, and the perfectives, by black points. Figures 6-9 zoom in on several subparts of the graphs focusing on particular verb groupings. The coordinates along the three axes (D1, D2, D3, which stand for the respective dimensions) are used to refer to the regions of the graph.

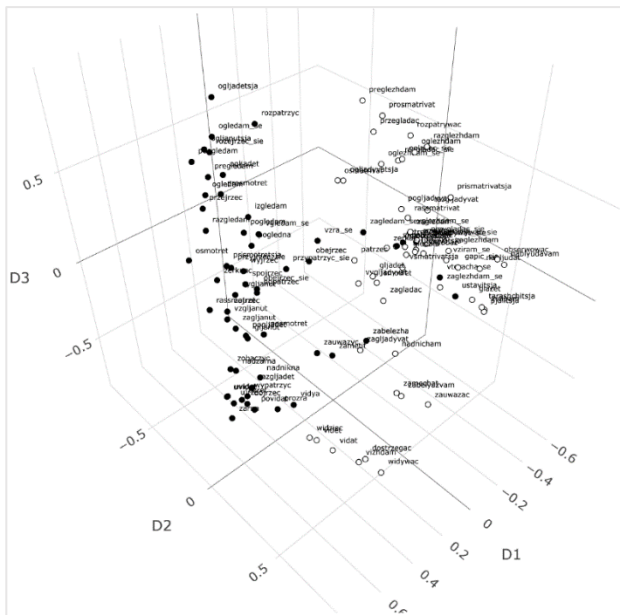Figure 5. MDS graph for the sight verbs under analysis

Figure 5 shows that aspect is one of the major distinctions structuring the verbs under analysis. Apart from a few exceptions to be discussed below, the perfective verbs are found on the left side and the imperfective verbs on the right side of Figure 5. The higher degree of correspondence between verbs with the same aspectual value was already observed in the previous sections, and the picture in Figure 5 further confirms its overarching role in structuring the sight verbs. As shown in the following discussion, with very few exceptions, the semantic groupings found on the perfective side of the graph are mirrored on the imperfective side.

Figure 6. Experience verbs and verbs similar to them

21
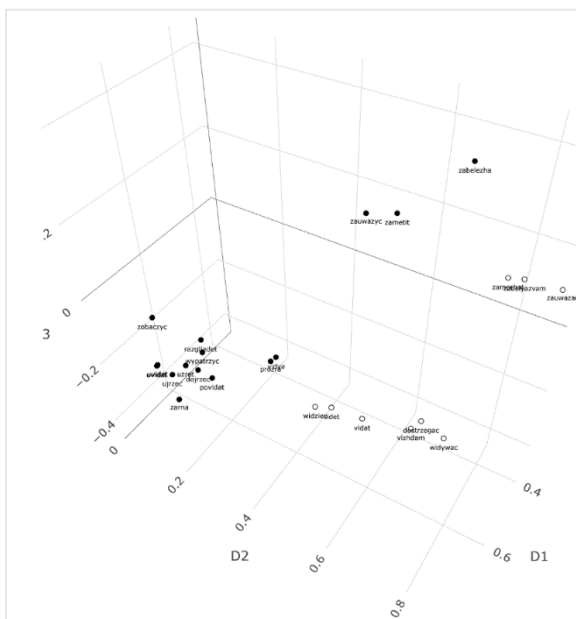
One of the smaller groupings, shown on Figure 6, can be found in the part of the graph where both D1 and D2 have positive values (with the exclusion of the verbs *osmatrivat* and *rozpatrywac*, found in the extreme positive values of D3). In the lower part of Figure 6, one finds basic experience verbs, such as B *vizhdam*, as well as their more marginal quasisynonyms, e.g., *uvidat'*, almost indistinguishable from *uvidet'* on the plot. This group also includes P *wypatrzyć* and R *razgljadet'*, which have the roots of activity verbs but denote the fact of perception specifying that it was accompanied by difficulties. In the top part, the 'notice' verbs of the three languages are situated. As shown in Figure 3 in section 5, they are somewhat closer to experience than to activity verbs but generally are separate from both basic verb types. Note that these small groupings are further subdivided in terms of aspect. These verbs are excluded from the figures discussed below.

Figure 7. Verbs of complete perception
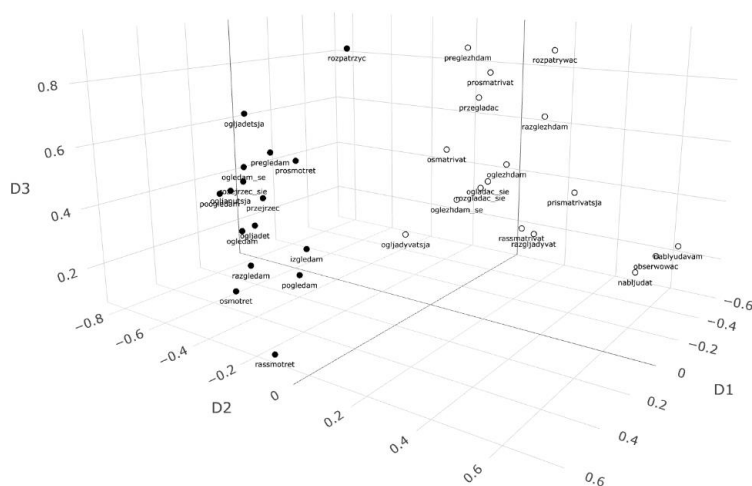
Let us now consider the top part of the graph (with the values of D3 above 0.06), shown in Figure 7. The two larger clouds of perfective and imperfective verbs found in the values of D2 lower than 0.4 mostly consist of prefixed verbs which, from a derivational perspective, convey the idea of looking through (e.g., R *prosmotret'*) or looking around (e.g., B *oglezhdam*). Reflexive verbs, describing looking around oneself, are more tightly grouped in the two aspectual clouds and arguably more directly express the spatial schema associated with the prefix. Non-reflexive verbs found here are more scattered, but the common semantic ground for their grouping might be that of complete perception of the observed object. The verbs of this type in the three languages are less closely grouped together than the verbs of looking through something, so they might be more semantically differentiated across the three languages. A small separate grouping in the higher values of D2 includes the 'observe' verbs.

Moving further down the graph, in the extreme negative values of D1 we mostly find imperfective verbs of attentive perception, such as B *vglezdam se*, which form a tight group, as already discussed in section 6 and clearly seen in Figure 5. A smaller subgrouping in the higher values of D2 includes some

of the derogative verbs of this semantic type (e.g., Russian *pjalit'sja*), although the Russian verb *ustavit'sja* and the Polish verb *gapić się* are closer to the stylistically neutral verbs of this type. In this part of the graph, we also observe a less consistent grouping of verbs in terms of aspect. Among imperfective verbs of attentive perception, we find some inchoative perfective verbs that describe entry into this type of perception.

Figure 8. Verbs of attentive perception

Figure 9. Activity verbs of the bottom part of the MDS-graph

The remaining verbs from the bottom part of the graph are shown in Figure 9. On the perfective side, we see the remaining verbs of attentive perception, e.g., P *przypatrzyć się*. Below them are found verbs of brief gaze together with basic perfective activity verbs, and at the very bottom of the graph, verbs of looking from behind or over an obstacle, such as B *nadzarna*, which describe essentially the same spatial configuration, viewed from two different perspectives. This order is mirrored on the imperfective side of the graph. In particular, imperfective basic activity verbs are found at around the same values of

D3 as their perfective counterparts. Brief perception verbs have no counterparts on the imperfective side.

Based on the data of these three languages, the following tentative semantic classification of sight verbs in Slavic may be proposed: basic experience verbs and their (quasi-)synonyms; 'notice' verbs; activity verbs: complete perception verbs; attentive perception verbs (mostly imperfective); 'observe' verbs; temporally limited perception verbs, including basic ones (perfective); verbs of looking from behind or over an obstacle; imperfective basic activity verbs.

Apart from this classification, the analysis of the structure of sight verbs confirmed many of the observations made in the previous sections on the basis of more specific parameters. They are discussed in the next section among the overall findings of this study.

8. Conclusions

This paper presented a quantitative exploration of semantic relations in the domain of sight between verbs in three Slavic languages – Bulgarian, Polish and Russian. Using the frequency of correspondences between a wide range of sight verbs in parallel texts, the study examined the pairs of verbs with the highest degree of correspondence between languages, the degree of correspondence of sight verbs of one language to basic sight verbs in the other languages, the degree to which the uses of a verb are spread across the uses of non-basic verbs in the other languages, and the groupings of all the verbs of sight under analysis.

At the basis of this study is the premise that correspondences in parallel texts can be used to compare the verbs' ranges of use and thereby establish the degree of their semantic similarity. The patterns of correspondence are not uniform across verbs – some verbs more neatly correspond to just one particular verb in another language, the contexts with some verbs are more evenly scattered across a number of verbs in another language. Despite the complexity of these patterns of correspondence, some semantic groupings of sight verbs can be discerned.

The study has shown the ever-present relevance of the aspectual distinction for the structuring of Slavic verbs, which may reflect the general patterns of aspectual usage not restricted just to sight verbs. Smaller semantic groupings identified in the domain of sight in most cases cut across aspectual distinctions and in others, they are restricted to a single aspect. Many of these groupings are associated with specific (groups of) prefixes.

The distinction between experience and activity verbs, traditionally drawn for basic sight verbs, also plays an important role in structuring the domain. Some non-basic verbs show more semantic similarity to experience verbs, some to activity verbs, and some others to neither of the two but there are no verbs clearly lying in between the two poles. Interestingly, the verbs that contradict this generalization are the basic verbs themselves. This fact can be at least partially explained by the cross-linguistic variation between experience and activity verbs in discourse uses, which are likely to be more extensive for basic sight verbs.

Non-basic verbs that are semantically close to basic activity verbs are more numerous that those that are close to basic experience verbs. This might be due to the fact that there are more manner and spatial configurations as well as the differences in assessment for controlled actions of directing attention rather than to the uncontrolled perception denoted by experience verbs. The presence of many verbs in the activity domain makes it more likely to be subject to lexical renewal and fluctuations in the verbs' ranges of uses.

The likeliest rivals of basic verbs are expected not only to share many of their contexts with basic verbs of other languages but also be semantically general enough to correspond to a large number of non-basic sight verbs. The latter ability can be assessed on the basis of the evenness of the spread of the verbs' uses across non-basic verbs. This property was shown to be positively correlated with the proportion of correspondences to basic activity verbs. The third parameter that makes a verb similar to basic, and independent of the two already discussed, is frequency. Thus, verbs with a high proportion of correspondences to basic verbs of the other languages, with a more even distribution of uses across non-basic verbs, and with high frequency are closer to basic verbs of their own language and can be regarded as their potential substitutes.

Frequency was found to positively correlate with the degree of mutual correspondence between the best-matching pairs of verbs. The tentative explanation of this finding is that the correspondence between more frequent verbs is likely to be less sensitive to semantic shifts and extensions that affect one of them.

Appendix

For each pair, the left column gives the language with a larger number of verbs.

Table 1A. Bulgarian-Polish pairs of verbs with the highest degree of correspondence based on the Dice-coefficient

Polish verbs	Aspect	Bulgarian verbs	Aspect	Dice coefficient
zauwazyc	pf	zabelezha	pf	0.72
obserwowac	ipf	nabljudavam	ipf	0.64
widziec	ipf	vizhdam	ipf	0.62
rozpatrywac	ipf	razglezhdam	ipf	0.59
rozejrzec sie	pf	ogledam se	pf	0.54
zobaczyc	pf	vidja	pf	0.53
spojrzec	pf	pogledna	pf	0.51
gapic sie	ipf	zjapam	ipf	0.50
przejrzec	pf	pregledam	pf	0.50
przegladac	ipf	preglezhdam	ipf	0.44
patrzec	ipf	gledam	ipf	0.40
rozgladac sie	ipf	oglezhdam se	ipf	0.35

ogladac	ipf	gledam	ipf	0.33
zauwazac	ipf	zabeljazvam	ipf	0.31
rozejrzec sie	pf	ogledam	pf	0.28
wpatrywac sie	ipf	vziram se	ipf	0.22
ogladac sie	ipf	oglezhdam se	ipf	0.20
zagladac	ipf	nadnicham	ipf	0.20
rozgladac sie	ipf	oglezhdam	ipf	0.18
przyjrzec sie	pf	razgledam	pf	0.17
zajrzec	pf	nadnikna	pf	0.16
rozpatrzyc	pf	razgledam	pf	0.15
patrzyc	ipf	gledam	ipf	0.14
popatrzec	pf	pogledna	pf	0.14
zerkac	ipf	nadnicham	ipf	0.11
dostrzegac	ipf	zabeljazvam	ipf	0.11
wyjrzec	pf	nadnikna	pf	0.10
rozejrzec sie	pf	poogledam	pf	0.08
wpatrywac sie	ipf	zagledam	pf	0.08
obejrzec	pf	gledam	ipf	0.08
przygladac sie	ipf	nabljudavam	ipf	0.08
obejrzec	pf	izgledam	pf	0.08
wpatrywac sie	ipf	vtrencha se	pf	0.08
wpatrzyc sie	pf	vtoracha se	pf	0.07
przyjrzec sie	pf	vgledam se	pf	0.07
ujrzec	pf	zarna	pf	0.06
zerknac	pf	nadnikna	pf	0.06
zapatrzyc sie	pf	zagledam se	pf	0.05
przypatrywac sie	ipf	vglezhdam se	ipf	0.05
zerknac	pf	nadzarna	pf	0.05
spogladac	ipf	vziram se	ipf	0.05
obejrzec sie	pf	ogledam se	pf	0.04
ogladac sie	ipf	zaglezhdam	ipf	0.04
widywac	ipf	vizhdam	ipf	0.04
przypatrzyc sie	pf	vzra se	pf	0.03
przejrzec	pf	prozra	pf	0.03
obejrzec	pf	pogledam	pf	0.03
ogladac sie	ipf	zaglezhdam se	ipf	0.02
wypatrzyc	pf	zarna	pf	0.02
dojrzec	pf	zarna	pf	0.01

Table 2A. Bulgarian-Polish pairs of verbs with the highest degree of correspondence based on the Dice-coefficient

Russian verbs	Aspect	Bulgarian verbs	Aspect	Dice coefficient
zametit	pf	zabelezha	pf	0.71
videt	ipf	vizhdam	ipf	0.63
smotret	ipf	gledam	ipf	0.56

nabljudat	ipf	nablyudavam	ipf	0.52
ogljanutsja	pf	ogledam se	pf	0.48
prosmatrivat	ipf	preglezhdam	ipf	0.43
zamechat	ipf	zabelyazvam	ipf	0.43
prosmotret	pf	pregledam	pf	0.39
uvidet	pf	vidya	pf	0.39
posmotret	pf	pogledna	pf	0.36
pjalitsja	ipf	zyapam	ipf	0.34
rassmatrivat	ipf	razglezhdam	ipf	0.28
ogljadetsja	pf	ogledam se	pf	0.28
ogljadyvatsja	ipf	oglezhdam	ipf	0.26
zagljadyvat	ipf	nadnicham	ipf	0.25
vzgljanut	pf	pogledna	pf	0.25
zagljanut	pf	nadnikna	pf	0.22
osmotret	pf	pregledam	pf	0.21
ustavitsja	ipf	zyapam	ipf	0.20
vygljadyvat	ipf	nadnicham	ipf	0.18
rassmotret	pf	razgledam	pf	0.18
osmatrivat	ipf	preglezhdam	ipf	0.16
prismotretsja	pf	vgledam se	pf	0.14
razgljadyvat	ipf	razglezhdam	ipf	0.12
vygljanut	pf	nadnikna	pf	0.11
ustavitsja	pf	vtrencha se	pf	0.10
vsmatrivatsja	ipf	vglezhdam se	ipf	0.10
gljadet	pf	zagledam se	pf	0.09
glazet	ipf	zyapam	ipf	0.09
vgljadyvatsja	ipf	vziram se	ipf	0.09
zagljanut	pf	nadzarna	pf	0.08
ogljadet	pf	ogledam	pf	0.08
prismatrivatsja	ipf	vglezhdam se	ipf	0.08
vgljadyvatsja	pf	vzra se	pf	0.07
tarashchitsja	ipf	zyapam	ipf	0.07
glazet	ipf	zaglezhdam	ipf	0.05
razgljadet	pf	ogledam	pf	0.05
ogljadetsja	pf	poogledam	pf	0.05
gljanut	pf	pogledna	pf	0.03
pogljadet	pf	izgledam	pf	0.03
pogljadyvat	ipf	oglezhdam se	ipf	0.03
uzret	pf	zarna	pf	0.03
vzirat	pf	vtoracha se	pf	0.02
zagljadyvat	ipf	zaglezhdam se	ipf	0.02
ljubovatsja	ipf	razglezhdam	ipf	0.02
pogljadet	pf	pogledam	pf	0.02
ustavitsja	pf	zagledam	pf	0.02
razgljadet	pf	prozra	pf	0.02

vidat	ipf	vizhdam	ipf	0.01
povidat	pf	vidya	pf	0.01
uvidat	pf	zarna	pf	0.00

Table 3A. Polish-Russian pairs of verbs with the highest degree of correspondence based on the Dice-coefficient

Russian verbs	Aspect	Polish verbs	Aspect	Dice coefficient
videt	ipf	widziec	ipf	0.88
zametit	pf	zauwazyc	pf	0.74
vygljanut	pf	wyjrzec	pf	0.60
uvidet	pf	zobaczyc	pf	0.57
prosmatrivat	ipf	przegladac	ipf	0.55
nabljudat	ipf	obserwowac	ipf	0.53
prosmotret	pf	przejrzec	pf	0.47
ogljadyvatsja	ipf	ogladac sie	ipf	0.46
zagljanut	pf	zajrzec	pf	0.46
zagljadyvat	ipf	zagladac	ipf	0.45
smotret	ipf	patrzec	ipf	0.40
pjalitsja	ipf	gapic sie	ipf	0.36
ogljanutsja	pf	rozejrzec sie	pf	0.36
posmotret	pf	spojrzec	pf	0.35
ustavitsja	pf	gapic sie	ipf	0.30
ogljanutsja	pf	obejrzec sie	pf	0.28
zamechat	ipf	zauwazac	ipf	0.28
ogljadetsja	pf	rozejrzec sie	pf	0.25
smotret	ipf	ogladac	ipf	0.25
vzgljanut	pf	spojrzec	pf	0.23
rassmotret	pf	przyjrzec sie	pf	0.18
ogljadyvatsja	ipf	rozgladac sie	ipf	0.18
pogljadyvat	ipf	zerkac	ipf	0.16
zamechat	ipf	dostrzegac	ipf	0.14
smotret	ipf	patrzyc	ipf	0.14
rassmatrivat	ipf	przygladac sie	ipf	0.10
gljadet	ipf	spogladac	ipf	0.10
osmotret	pf	obejrzec	pf	0.10
razgljadet	pf	przyjrzec sie	pf	0.10
razgljadyvat	ipf	przygladac sie	ipf	0.09
vgljadyvatsja	ipf	wpatrywac sie	ipf	0.09
prismotretsja	pf	przyjrzec sie	pf	0.09
rassmotret	pf	rozpatrzyc	pf	0.09
vsmatrivatsja	ipf	wpatrywac sie	ipf	0.08
rassmatrivat	ipf	rozpatrywac	ipf	0.08
tarashchitsja	ipf	gapic sie	ipf	0.07
glazet	ipf	gapic sie	ipf	0.06
vzgljanut	pf	zerknac	pf	0.06

razgljadet	pf	dojrzec	pf	0.05
vygljadyvat	pf	wyjrzec	pf	0.05
vgljadyvatsja	ipf	przypatrywac sie	ipf	0.05
gljanut	pf	popatrzec	pf	0.04
ogljadet	pf	rozejrzec sie	pf	0.04
pogljadet	pf	popatrzec	pf	0.03
ustavitsja	pf	zapatrzyc sie	pf	0.03
prismotretsja	pf	przypatrzyc sie	pf	0.03
prismatrivatsja	ipf	przygladac sie	ipf	0.03
osmatrivat	ipf	przygladac sie	ipf	0.03
ljubovatsja	ipf	wpatrzyc sie	pf	0.02
uzret	pf	ujrzec	pf	0.02
vzirat	ipf	wpatrywac sie	ipf	0.02
vidat	ipf	widywac	ipf	0.02
uvidat	pf	ujrzec	pf	0.02
ogljadet	pf	wypatrzyc	pf	0.01
povidat	pf	zobaczyc	pf	0.01

References

Dickey, Stephen. 2000. Parameters of Slavic Aspect. A cognitive approach. Stanford, CACSLI Publications.

Divjak, Dagmar. 2015. Exploring the grammar of perception. A case study using data from Russian. Functions of Language 22.1, 44–68.

Evert, Stefan. 2009. Corpora and collocations. In: Anke Lüdeling & Merja Kytö (Eds.), Volume 2: An International Handbook, pp. 1212–1248. Berlin, New York: De Gruyter Mouton. DOI: 10.1515/9783110213881.2.1212

Georgakopoulos, Th., E. Grossman, D. Nikolaev & S. Polis. 2021. Universal and macro-areal patterns in the lexicon: A case-study in the perception-cognition domain. Linguistic Typology, 26(2), 2021–2088.

Hausser, Jean, and Korbinian Strimmer. 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. J. Mach. Learn. Res., 10, 1469–1484. Available online: https://jmlr.csail.mit.edu/papers/v10/hausser09a.html

Hilpert, Martin; Correia Saavedra, David. 2017. Why are grammatical elements more evenly dispersed than lexical elements? Assessing the roles of text frequency and semantic generality. Corpora, 12(3), 369–392. DOI: 10.3366/cor.2017.0125

Knjazev, Y. P. 2007. Грамматическая семантика: Русский язык в типологической перспективе. Москва.

Levshina, Natalia. 2016. Why we need a token-based typology: A case study of analytic and lexical causatives in fifteen European languages. Folia Linguistica, 50(2), 507–542. DOI: 10.1515/flin-2016-0019.

Levshina, Natalia. 2017. Online film subtitles as a corpus: An n-gram approach. Corpora, 12(3), 311–338. DOI: 10.3366/cor.2017.0123.

Levshina, Natalia. 2015. How to do Linguistics with R: Data exploration and statistical analysis. Amsterdam, Philadelphia: John Benjamins.

Mair, P., Groenen, P. J. F., De Leeuw, J. 2022. More on multidimensional scaling in R: smacof version 2. Journal of Statistical Software, 102(10), 1–47.

Nesset, Tore, Laura A. Janda, Julia Kuznetsova, OX'PK Lyashevskaya, Anastasia Makarova, Svetlana Sokolova. 2008. Why poslushat', but uslyshat'? Poljarnyj Vestnik 11. 38-46.

Nesset, Tore. 2010. Is the Choice of Prefix Arbitrary? Aspectual Prefixation and Russian Verbs of Perception. Slavic and East European Journal, 54(4), 666–689.

Norcliffe, Elisabeth, Asifa Majid. 2024. Verbs of perception: A quantitative typological study. Language, Vol. 100(1), 81-123.

Ovsjannikova – Kukova. 2022. Овсянникова, М., & Кукова, Х. Предикаты восприятия в болгарском и русском языках и их сентенциальные актанты. In: Коева, С. & Иванова, Е. (Eds.), Онтология на ситуациите за състояние – лингвистично моделиране, pp. 304–357. София.

Ovsjannikova, M. A. 2019. "Sintaksičeskoe kodirovanie èksperiencial′nych situacij v russkom jazyke: sopostavlenie grupp glagolov vosprijatija, myšlenija i èmocij." *Voprosy jazykoznanija* 6.

Ovsjannikova, Kukova. 2023. Clausal Complements of Non-Agentive Visual Perception Verbs in Bulgarian and Russian. *Slavia* XCII (2). P. 145–169.

Padučeva, E. V. 2004. *Dinamičeskie modeli v semantike leksiki.* Moscow: Jazyki slavjanskoj kul′tury.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Rosen, A., M. Vavřín, A. J. Zasina. 2022. InterCorp, Release 15 of 11 November 2022. Institute of the Czech National Corpus, Charles University. Available at: http://www.korpus.cz

Rosen, Alexandr. 2023. The InterCorp Parallel Corpus with a Uniform Annotation for All Languages. Journal of Linguistics/Jazykovedný časopis, 74(1), 254–267.

San Roque, Lila, Elisabeth Norcliffe, Kobin H. Kendrick, Majid, Asifa et al. 2015. Vision verbs dominate in conversation across cultures. In: Aikhenvald, Alexandra Y., and Anne Storch (eds.), Perception and Cognition in Language and Culture, 45–72. Brill. DOI: 10.1163/9789004280205_004.

Roque, Lila San, Kendrick, Kobin H., Norcliffe, Elisabeth, Majid, Asifa. 2018. Universal meaning extensions of perception verbs are grounded in interaction. Cognitive Linguistics, 371-406. https://doi.org/10.1515/cog-2017-0034

Singhal, Amit. 2001. Modern Information Retrieval: A Brief Overview. IEEE Data Engineering Bulletin, 24(4), 35–43.

Stoll S., Mazara J., Bickel B. (2017) The acquisition of polysynthetic verb forms in Chintang // Fortescue M., Mithun M., Evans N. (Eds.) The Oxford Handbook of Polysynthesis. Oxford: Oxford University Press, 495–514.

Sweetser. 1990. From Etymology to Pragmatics. Metaphorical and cultural aspects of semantic structure. Cambridge: Cambridge University Press.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. Journal of Child Language, 31(1), 61–99. https://doi.org/10.1017/S0305000903005881&#8203

Viberg, Åke. 1984. The verbs of perception: a typological study. *Linguistics* 21.1, 123–162.

Viberg, Åke. 2001. Verbs of perception. In: M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), Language Typology and Language Universals: An International Handbook, pp. 1294–1309. Berlin: Walter de Gruyter.

von Waldenfels, Ruprecht. 2012. Aspect in the imperative across Slavic—a corpus driven pilot study. In: Atle Grønn & Anna Pazelskaya, eds., The Russian Verb. Oslo Studies in Language 4, 141–154.

Wälchli, B. 2016. Non-specific, specific and obscured perception verbs in Baltic languages. Baltic Linguistics, 7, 53–135.