Slovo. Journal of Slavic Languages, Literatures and Cultures
No. 65, 2025

Ukrainian Indefinite Pronouns and Language Typology

Original research article

Tore Nesset and Yuliia Palii
Department of Language and Culture, UiT The Arctic University of Norway

Tore.nesset@uit.no, yuliia.palii@uit.no

Abstract

The present article offers an empirical analysis of Ukrainian indefinite pronouns and adverbs based on data from the GRAC corpus. The proposed analysis has ramifications for Ukrainian linguistics, Slavic linguistics, and language typology. With regard to Ukrainian linguistics, we identify substantial frequency differences and suggest distinguishing between a "core" system including the indefiniteness markers *de-*, *-s'* and *bud'-*, and three "peripheral" markers, viz. *-nebud'*, *aby-*, and *kazna*. From the perspective of Slavic linguistics, the proposed analysis facilitates comparison with other Slavic languages, such as Polish and Russian. Our analysis pinpoints a number of similarities across Ukrainian, Polish and Russian, but also demonstrates that Ukrainian has a distinct system that merits investigation in its own right. For language typology, the analysis we propose shows how frequency information can be integrated in semantic maps, which arguably makes semantic maps a more powerful tool for cross-linguistic comparison.

Introduction

Indefinite pronouns have attracted considerable attention in Slavic linguistics, and in his seminal typological study, Haspelmath (1997) includes thorough analyses of *inter alia* Polish and Russian. Haspelmath (1997) does not analyze Ukrainian, which has received less attention in scholarly literature than other Slavic languages (but see Fisun 2019, 2022). In this article, we present an analysis of Ukrainian indefinite pronouns and adverbs based on data from the GRAC corpus (Shvedova et al. 2017–22).

In Slavic languages, indefinite pronouns and adverbs are typically formed by adding what we may call an indefiniteness marker to a question word (a pronoun like Ukrainian *xto* 'who' or an adverb like Ukrainian *de* 'where'). For the purposes of the present article, we will consider pronouns and adverbs based on the following indefiniteness markers: *de-, -s', bud'-, -nebud', aby-,* and *kazna.* As demonstrated by Fisun (2019, 2022), more markers are attested, but these fall outside the scope of the present article, which focuses on the most conventionalized and prototypical indefinite pronouns and adverbs in Ukrainian.

Our contribution can be summarized as follows. First, with regard to Ukrainian linguistics, we offer a semantic map for the relevant pronouns and adverbs and analyze the frequency distributions of the words under scrutiny. On the basis of frequency data, we propose a distinction between a "core" system and what we refer to as "peripheral" markers. Second, from the perspective of Slavic linguistics, we offer a comparison of Ukrainian with Russian and Polish, showing that the Ukrainian system is different from both neighboring languages. Third, when it comes to language typology, we argue that semantic maps can yield more insightful analyses if frequency information is included in the maps.

After a brief discussion of semantic maps in section 2, we present the frequency data for Ukrainian and carry out a correspondence analysis of these data in section 3. Section 4 is an analysis of the Ukrainian "core" system, which is compared to Polish and Russian in section 5. In section 6, we explore "peripheral" markers, before we show how frequency data can be incorporated in semantic maps in section 7. Section 8 sums up our findings.

1. Semantic maps for indefinite pronouns and adverbs

Haspelmath (1997) takes nine broad categories as the point of departure for his typological survey of indefinite pronouns, which are illustrated with Russian examples (from Haspelmath 1997:273–275):

(1) Specific known:
Nam nado pogovorit' s toboj koe o čem.
'You and I have to talk about something.'

(2) Specific unknown:
Kto-to postučal v dver'.
'Someone knocked on the door.'
(3) Irrealis non-specific:
My vstretimsja gde-nibud'
'We'll meet somewhere.'
(4) Question:
Zvonil mne kto-nibud'?
'Did anyone call me?'
(5) Conditional:
Esli čto-nibud' slučitsja, ja skažu mame.'
'If anything happens, I'll tell mom.'
(6) Comparative:
Zdes' prijatnee žit' čem gde-libo v mire.
'It is more pleasant to live here than anywhere in the world.'
(7) Indirect negation:
bez kakoj-libo pomošči
'without any help'
(8) Free choice:
Ty možeš' kupit' ljubuju knigu.
'You may buy any book.'
(9) Direct negation:
Nikogda ja ne zabudu tebja.
'I will never forget you.'

While, as shown in the copious literature on Russian indefinite pronouns (e.g., Dahl 1970, Padučeva 1985: 209–220, Padučeva 2011, Padučeva 2018, Tatevosov 2002), more fine-grained distinctions can be pinpointed, these broad categories represent a good starting point for comparison of languages and will be adopted in this article. We note, though, that Haspelmath's categories represent a heterogeneous set. Some of the categories describe the referential properties of the relevant lexical items. For instance, the distinction between (1) and (2) has to do with whether the referent is known to the speaker or not. In a similar vein, (1) and (2) represent the difference between specific and non-specific reference. Other categories describe contextual features that license certain indefinite pronouns. Examples of this type are questions in (4) and conditional clauses in (5).

In the following, we focus on categories (1)–(8). Category (9), direct negation, behaves the same way in Ukrainian, Polish and Russian, and is therefore not of primary interest when comparing these languages.

A semantic map may be defined as "a way to visually represent the interrelationships between meanings expressed in languages" (Georgakopoulos and Polis 2018:1). Haspelmath (1997) distributes the nine categories so that closely related categories are close to each other on the map. Lines indicate the categories covered by certain forms. The semantic map in Figure 1, adapted from Haspelmath (1997:271), describes the distribution of the three Polish markers *-s'*, *-kolwiek,* and *ni-*. Hasplemath (1997:273)

also includes a semantic map for Russian. As shown in Figure 2, this map involves seven markers: *koe-, -to, -nibud', -libo, ni-, ljuboj,* and *ugodno.*

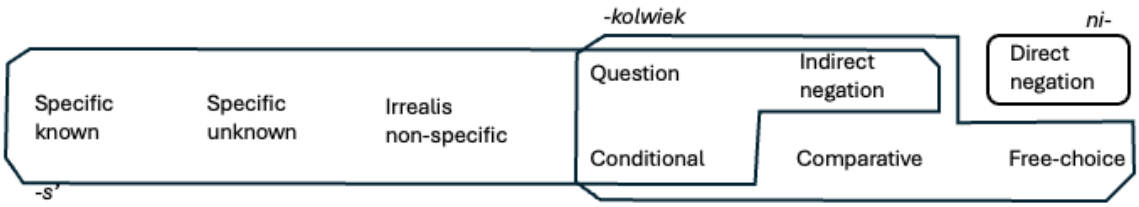


Figure 2: Semantic map for Polish (adapted from Haspelmath 1997:271)

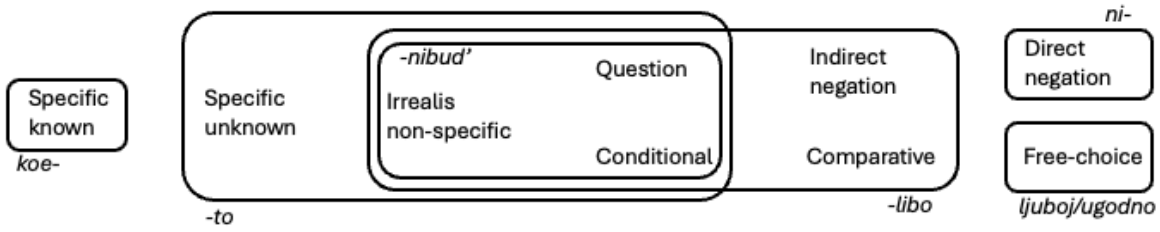


Figure 3: Semantic map for Russian (adapted from Haspelmath 1997:273)

Haspelmath does not provide a semantic map for Ukrainian. It is therefore necessary to present the relevant indefiniteness markers and their frequency distribution in section 3, before we turn to more detailed analyses in the following sections.

2. Frequency data and correspondence analysis

We consider six indefiniteness markers (*de-, -s', bud'-, -nebud', aby-,* and *kazna*) in combination with seven widely used question words: *xto* 'who', *de* 'where', *jakyj* 'which', *koly* 'when', *jak* 'how', *ščo* 'what' and *kudy* 'whereto'. Table 1 displays the frequencies of all combinations of indefiniteness markers and question words in the GRAC corpus (Shvedova et al. 2017–2022).[141]

The frequency differences are substantial. As shown in Figure 3, the three markers *-s'* (66.7%), *de-* (19.0%) and *bud'-* (12.2%) together cover 97.9% of the corpus examples in Table 1. It therefore makes sense to consider these three markers the "core" system of Ukrainian. We refer to the remaining markers as "peripheral", since they constitute only 2% of the examples in Table 1.

	de-	*-s'*	*bud'-*	*-nebud'*	*aby*	*kazna*	Total
xto	73 424	387 081	19 927	10 609	367	55	491 463

[141] Corpus searches were carried out in May, 2025. We used data from version 18 of GRAC. The datasets analyzed in the present article are available at https://doi.org/10.18710/VESH4N.

de	0	218 927	3067	3440	105	58	225 597
jakyj	532 048	818 183	429 328	22 641	1141	111	1 803 452
koly	8469	197 534	7730	12 211	19	21	225 984
jak	0	181 626	355	1510	2286	47	185 824
ščo	122 438	736 190	12 063	20 458	777	1983	893 909
kudy	523	42 993	584	1612	138	65	45 915
Total	736 902	2 582 534	473 054	72 481	4833	2340	

Table 1. Frequency distributions for Ukrainian indefinite pronouns and adverbs

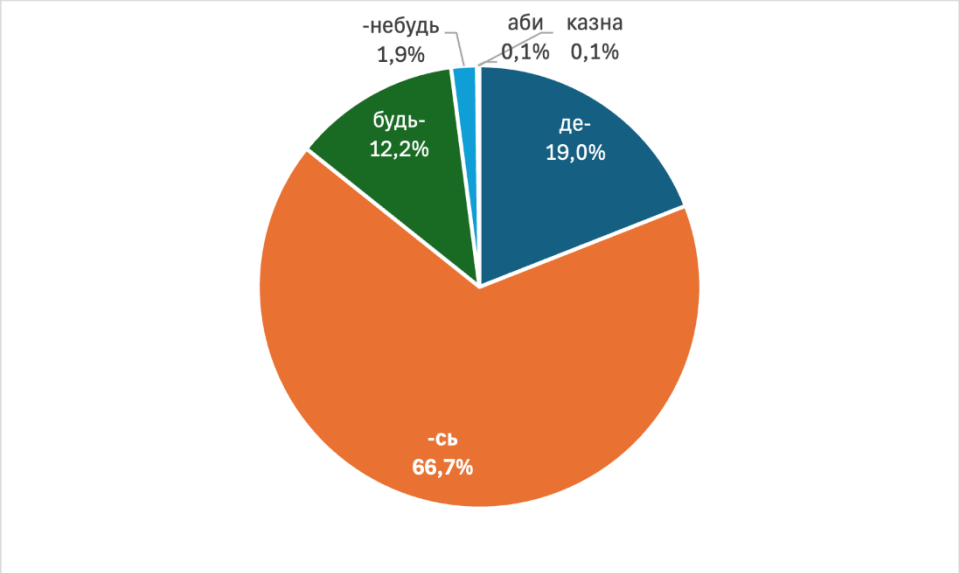


Figure 4: Relative frequencies of indefinite markers in Ukrainian (based on Table 1)

In order to get a better understanding of the relationships between indefiniteness markers and question words, we carried out a correspondence analysis of the "core" system on the basis of Table 1.[142] This model groups indefiniteness markers and question words in an n-dimensional space and maps out the two most important dimensions in the graph shown in Figure 4. In the figure, the indefiniteness markers are given in red, while black is used for question words. Notice that *de* is both an indefiniteness marker and a question word, and therefore occurs twice in the figure.

The horizontal dimension, which covers 97.2% of the variation, maps out the indefiniteness markers with *-s'* to the left and *de-* to the right (*bud'-* ended up outside the graph). Less important, but still relevant, is the vertical dimension, which covers

[142] We are grateful to Laura A. Janda for assistance with the correspondence analysis.

2.8% of the variation. This dimension shows that the question words behave differently. From top to bottom we get a scale ranging from substantival pronouns (*xto* 'who' and *ščo* 'what') via the adjectival pronoun *jakyj* 'which' to adverbs (*de* 'where', *koly* 'when', *jak* 'how', and *kudy* 'whereto'). The graph shows that *de-* is closest to pronouns, while *-s'* is located between pronouns and adverbs. Clearly, indefiniteness markers show different preferences for question words – and vice versa.

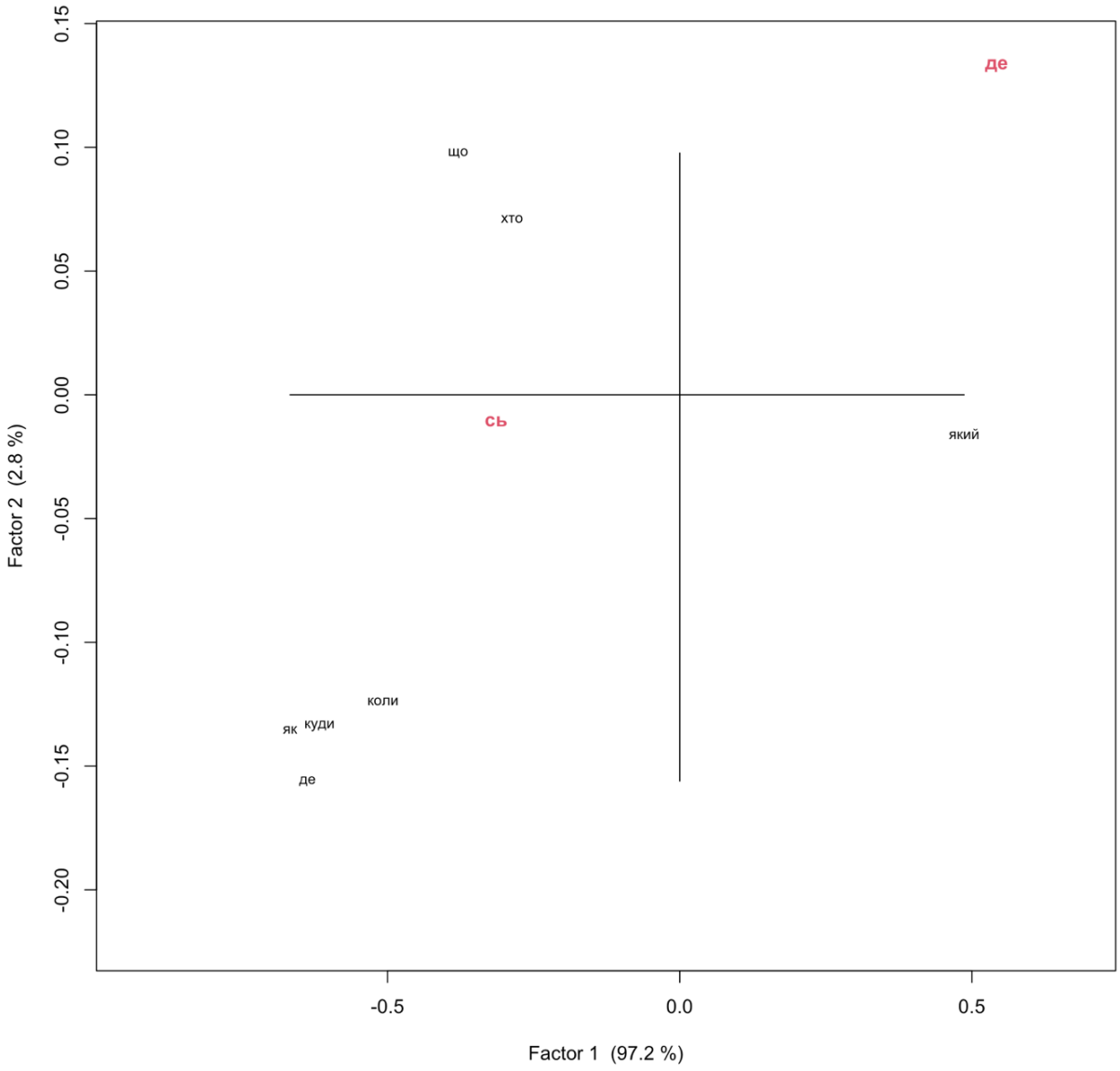


Figure 5. Correspondence analysis of the Ukrainian "core" system based on numbers from Table 1

The corpus data presented in this section clearly show that frequency matters. We return to this point in section 7, but first we must consider the distribution of the markers with regard to Hapelmath's categories. This is the topic of the next section.

3. The Ukrainian "core" system

3.1 The most frequent categories

We extracted a random sample of 600 examples from the GRAC corpus and annotated the data manually for the semantic categories in Table 2. As shown in Table 2, we excerpted 200 examples for each "core" marker. The data offers a good picture of the situation for the categories "specific known", "specific unknown", "irrealis non-specific" and "free choice".

	де	сь	будь	Total
specific_known	147	3	0	150
specific_unknown	53	148	0	199
irrealis_non_specific	0	40	0	40
free_choice	0	0	200	200
Other	0	9	0	9
Total	**200**	**200**	**200**	**600**

Table 2: The Ukrainian "core" system. Data sample from GRAC corpus

The relative frequencies are given in Figure 5. As shown, *de-* dominates for specific known, while *-s'* is the most widespread option for specific unknown and the only option for irrealis. For free choice, *bud'-* is used.

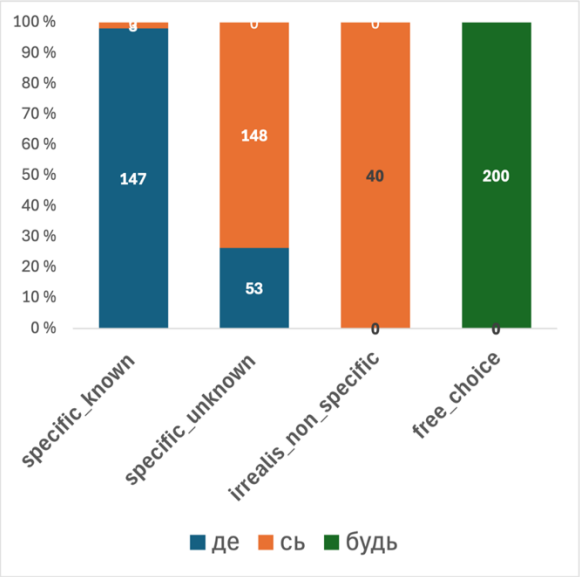


Figure 6: The Ukrainian "core" system. Data sample from GRAC corpus

Here are some illustrative examples

(10) Specific known:
Ty obicjav, rozumnyku, deščo rozpovisty.
'You promised, smart man, to tell me something.

(11) Specific unknown:
Inkoly bat'ky ne vkazujut' v anketax pro jakis' osoblyvosti dytyny.
'Sometimes parents do not include any characteristics of the child in the questionnaires.'

(12) Irrealis non-specific:
My ne možemo skazaty, pro jakis' perspektyvy, ale robota tryvaje.
'We can't say anything about any future options, but our work continues.'

(13) Free choice:
A čynovnyk, bud'-jakogo rangu, nese ščе j polityčnu vidpovidal'nist'.
'And an official, of any rank, also bears political responsibility.'

3.2 Questions

In order to investigate the distribution of indefiniteness markers in questions, we searched in the GRAC corpus for *deščo* or *ščos'* followed by "?" with up to five intervening words. The results are summarized in Table 3 and Figure 6, which demonstrate that *s'*- is the dominant marker that covers about 99% of the examples. Here is a relevant example *ščos'*:

(14) Jak možna z takoju ljudinoju pro ščos' govoryty?
'How could it be possible to talk to such a person about anything?'

	# examples
deščo	518
ščos'	37 631
Total	38 149

Table 3:* de-** ***vs.* -s' *in questions. Data from GRAC corpus

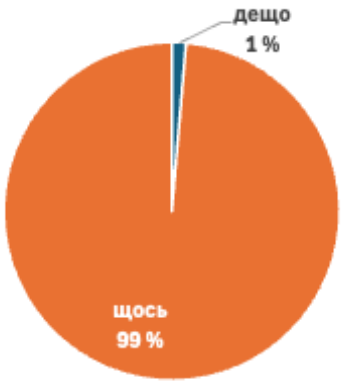


Figure 7:* de-** ***vs.* -s' *in questions. Data from GRAC corpus

3.3 Conditional clauses

We searched in GRAC for *jakščo*/*jakby* 'if' followed by *ščo*/*jakyj* with *de-*, *-s'*, *-nebud'* or *bud'* with up to five intervening words. The results are summarized in Table 4 and Figure 7. As shown, *-s'* dominates strongly with 84% of the examples. Here is an example with *-s'* (*jakas'* 'any kind' (nominative feminine singular)):

(15) Jakščo potribna jakas' dopomoga po xatnim spravam, ja prošu mamu abo babusju.'
'If I need any kind of help in my house, I ask my mother or grandmother,'

	# examples
de-	3062
-s'	43 514
-nebud'	1476
bud'-	3696
Total	51 748

Table 4: Conditional clauses in GRAC corpus

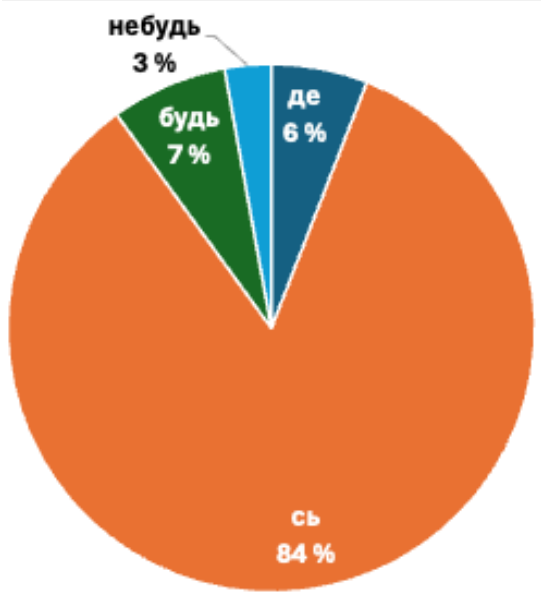


Figure 8: Conditional clauses in GRAC corpus

3.4 Indirect negation

Indirect negation concerns contexts where the preposition *bez* 'without' precedes an indefinite pronoun, as in the following example:

(16) Vin, bez bud'-jakogo sumnivu, zvažyvsja b na ostannje jakby buv sam.
'Without any doubt, he would have gone for the second option if he were alone.

The results, summarized in Table 5 and Figure 8, show that *bud'-* is the dominant marker for indirect negation, since it covers 89% of the examples.

	# examples
de-	280
-s'	1392
-nebud'	333
bud'-	16 354
Total	18 359

Table 5: Indirect negation in GRAC corpus

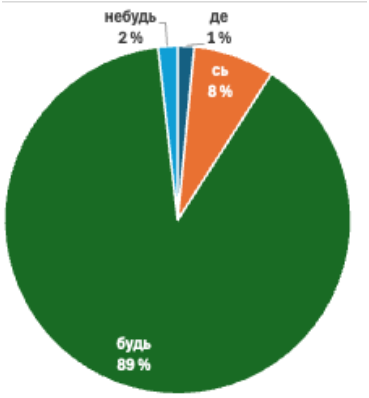


Figure 9: Indirect negation in GRAC corpus

3.5 Comparative

The comparative category involves *niž* 'than' followed by an indefinite pronoun or adverb:

(17) Hadaju, sportsmena ce spodobajet'sja bil'še, niž bud'-ščo z togo, ščo my možemo ïm zaproponuvaty.
'I think the athletes will like it more than anything we can offer them.'

We searched for *niž* 'than' followed by an indefinite pronoun or adverb and received the results summarized in Table 6 and Figure 9. As shown, *bud'* dominates with 83%, while *-s'* is attested in 11% of our examples.

	# examples

-s'	576
bud'-	2734
Total	3310

Table 6: Comparative in GRAC corpus

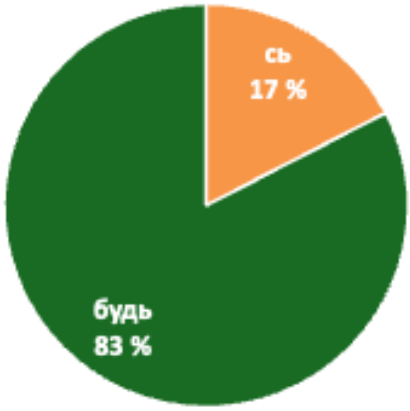
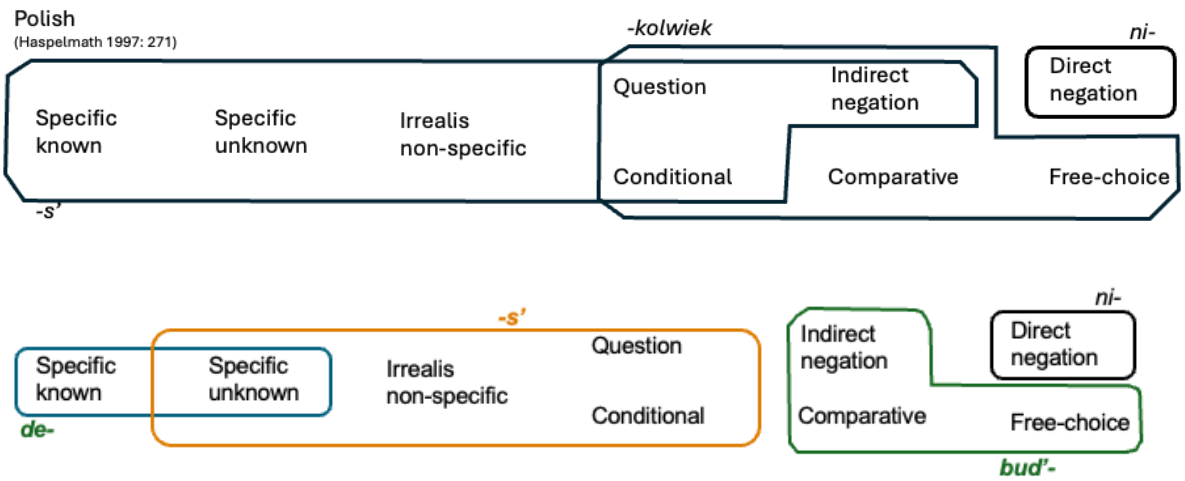


Figure 10: Comparative in GRAC corpus

4. Ukrainian "core" system vs. Polish and Russian

Based on the findings reported in section 4, we are in a position to draw a semantic map for Ukrainian and compare it with Haspelmath's maps for Polish and Russian (shown in section 2 above). For the convenience of the reader, we have juxtaposed the three maps in Figure 10, where Ukrainian is placed under Polish and above Russian.



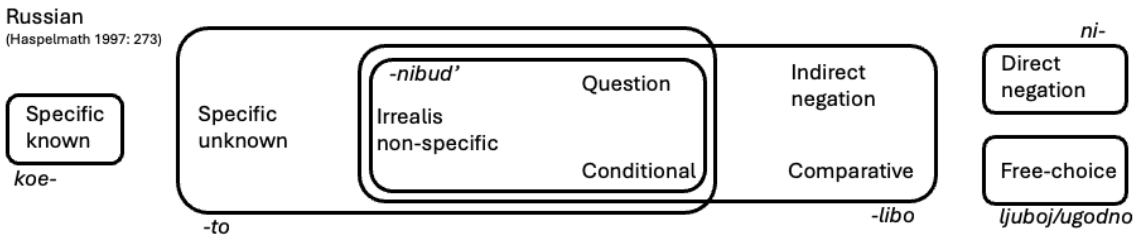


Figure 11: Comparison of semantic maps for Polish (top), Ukrainian (mid), and Russian (bottom)

As shown, in Ukrainian *de-* covers the two leftmost categories (specific known and specific unknown). Specific unknown also has *-s'*, which furthermore covers irrealis non-specific, question and conditional. In the rightmost portion of the Ukrainian map, *bud'-* is the dominant marker, covering indirect negation, comparative and free-choice. Arguably, we could have extended the area of *-s'* to include comparative, but we decided to limit ourselves to the dominating marker *bud'-* for this category. This simplification does not affect our line of argumentation in the following.

All three semantic maps have certain markers that gravitate towards the leftmost portion of the map, while others are found in the middle or in the rightmost portion of the map. Thus, Ukrainian *de-* has a distribution similar to Russian *koe-* in the leftmost part of the map. In the middle portion of the map, Ukrainian *-s'* is similar to Russian *-to*, while Polish *-s'* has a somewhat wider distribution than its Ukrainian counterpart. Ukrainian *bud'-* covers a smaller area than *kolwiek* in Polish, but a larger area than Russian *ljuboj/ugodno*, which are restricted to the free-choice category on Haspelmath's (1997:273) map.

Despite the similarities, it is clear that the Ukrainian "core" system is substantially different from both the Polish and the Russian systems. The Ukrainian system therefore merits study in its own right. In the following section, we expand our analysis of Ukrainian by adding "peripheral" markers to the map.

5. "Peripheral" markers in Ukrainian

In order to gather information about the "peripheral" markers (*-nebud'*, *aby-* and *kazna*), we extracted a random sample of 100 examples of each marker with *jakyj* and *ščo*. Since *kazna* turned out not to combine with *jakyj* in our dataset, for *kazna* we only have data for *ščo* with this marker.

For *-nebud'* all examples involved the category irrealis non-specific:

(18) –Zroby ščo-nebud', ščob ne vtopyvsja (Onlajn-ZMI, 2022)
'Do something, so he won't drown.'

The marker *aby-* was attested exclusively in the free-choice category:

(19) Pal'to deputatka obrala ne aby-jake, a vid italijs'kogo brendu Dolce&Gabbana. (Tablo ID, 2021)

‘The member of parlament didn’t choose just any jacket, but one of the Italian brand D&G.’

Kazna was also only attested in the free-choice category in our data sample:

(20) I ce vže ne texnologija, a kazna-ščo, tak pšenicju vyroščuvaty ne možna vzagali (Den’, 2000)
‘And this is not a good method, but goodness knows what, wheat cannot be grown like this.’

In Figure 11, we have added the “peripheral” markers to the semantic map for Ukrainian, proposed in section 5 above. The additions reinforce the conclusion reached in section 5. Although the Ukrainian system is similar to the systems of Polish and Russian, Ukrainian clearly has a distinct system that is different from those of the two neighboring languages under scrutiny in the present article.

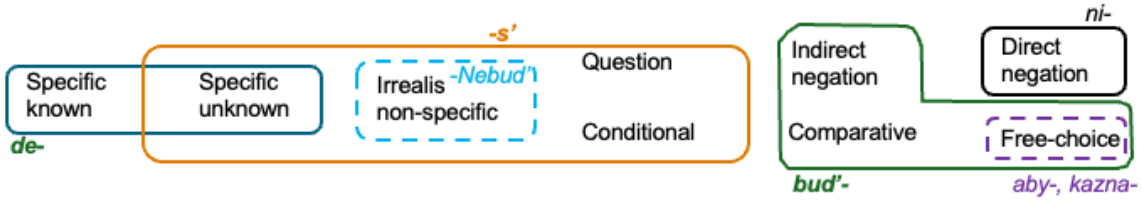


Figure 12: Semantic map for Ukrainian including “peripheral” markers

6. Semantic maps and frequency

In section 2, we demonstrated that frequency matters for indefinite pronouns and adverbs in Ukrainian. There are substantial frequency differences, and our correspondence analysis showed that different question words have different preferences for indefiniteness markers, and *vice versa*. It seems very likely that these different preferences are part of the mental grammars of native speakers, since there is ample evidence that native speakers are sensitive to frequency differences (e.g., Bybee 2007, Ellis et al. 2014)

Although semantic maps of the type used by Haspelmath (1997) represent a good starting point for cross-linguistic comparison and typological studies, the maps do not include frequency information. In other words, Haspelmath’s semantic maps distinguish between what is possible and what is impossible in a language but do not capture what is frequent or typical as opposed to what is infrequent or atypical.

We propose adding information about relative frequencies along two dimensions. With regard to the categories, which represent the meaning dimension, we suggest representing differences in relative frequency by means of different font sizes. For Ukrainian, we propose distinguishing between four levels:

(21) Categories (meaning): different font size
a. specific unknown: >60%
b. specific known, irrealis: 10–20%

c. question, free choice: 2%
d. other < 1%

When it comes to the markers, which represent the dimension of form, we propose using lines of different thickness to represent differences in relative frequency. For Ukrainian, it seems reasonable to adopt a tripartite distinction:

(22) Indefiniteness markers (form): different thickness of lines
a. *-s'*: > 60% (thick line)
b. de-, bud'-: 10–20% (thin line)
c. other: < 2% (dashed line)

In Figure 12, we provide a semantic map for Ukrainian with information about relative frequencies, as suggested above. Clearly, the proposed semantic map is somewhat impressionistic, although it is based on a quantitative analysis of frequency data. However, the map we propose does include more information than the maps of Haspelmath (1997), and we argue that the additional information is presented in a user-friendly way that makes it possible to get a good overview of the situation in the language.

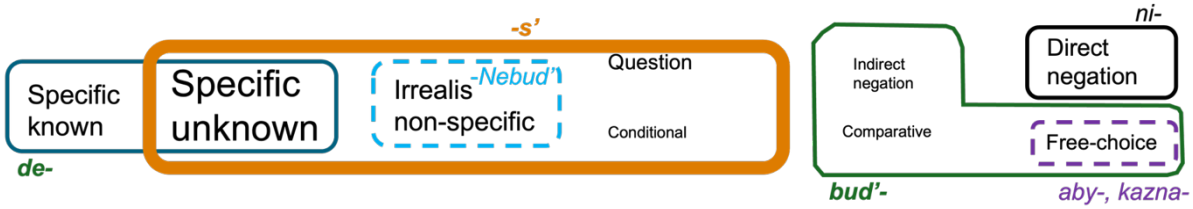


Figure 13: Semantic map for Ukrainian with integrated frequency information

We would like to emphasize that the map in Figure 12 includes information about *relative* frequencies. In other words, the differences reflected in the Figure concern the proportions of examples for each category or indefiniteness marker (cf. Table 1 in section 3 above).

We conclude that it is possible to include frequency information in semantic maps in a straightforward way. When we add frequency information about categories (meaning) and indefiniteness markers (form), the result is a diagram that is both informative and reader-friendly. We submit that this approach can be insightfully applied beyond indefiniteness pronouns and adverbs in Slavic.

7. Concluding remarks

In this article, we offer a corpus-based analysis of Ukrainian indefinite pronouns and adverbs which has implications for Ukrainian linguistics, Slavic linguistics, and language typology.

With regard to Ukrainian linguistics, our analysis shows that there are substantial frequency differences among the attested combinations of question words and

indefiniteness markers. A correspondence analysis clearly testifies to the different preferences of question words for indefiniteness markers – and *vice versa*. Based on frequency data, we propose distinguishing between a "core" system of frequent indefiniteness markers, and an additional system consisting of "peripheral" markers of lower frequency.

When it comes to Slavic linguistics, the proposed analysis of the Ukrainian data facilitates comparison with other Slavic languages, such as Polish and Russian, considered in the present article. Our comparison shows that although there are several similarities across the three languages under scrutiny, Ukrainian clearly posits a distinct system that merits analysis.

From a typological perspective, our analysis shows how information about relative frequencies can be incorporated into semantic maps in a straightforward way. We propose adding frequency information concerning both meaning (Haspelmath's nine categories) and form (indefiniteness markers).

The Ukrainian inventory of indefinite pronouns and adverbs is considerably larger than the pronouns and adverbs we have analyzed, as demonstrated by Fisun (2019, 2022). More research is needed in order to find out whether and how these pronouns and adverbs can be accommodated in the kind of analysis we propose. However, such an analysis is beyond the scope of this short article.

References

Bybee, Joan L. 2007. *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Dahl, Östen. 1970. "Some Notes on Indefinites". *Language* 46, 33–41.

Ellis, Nick C., O'Donnell, Matthew B. & Römer, Ute. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics* 25.1: 55–98.

Fisun, Roman. 2019. Zusammengesetzte Indefinitpronomen Im Ukrainischen. H. Macjuk, I. Mytnik, O. Novikova (eds.): *Language in society: semantics, syntax, pragmatics*. Warszawa-Lwów-Monachium, Lublin: Wydawnictwo UMCS, 85–99.

----------. 2022. Inventar' modifikatorov i istočniki grammatikalizacii sostavnyx neopredelennyx mestoimenij v slavjanskix jazykax. E.S. Uzeneva, O.V: Xavanova (eds.): *Slavjanskij mir. Obščnost' i mnogoznačie*. Moscow: Institut slavjanovedenija RAN, 80–85.

Georgakopoulos, Thanasis & Polis, Stéphane . 2018. "The semantic map model: State of the art and future avenues for linguistic research". *Lang Linguist Compass*. 12:e12270. https://doi.org/10.1111/lnc3.12270

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Padučeva, Elena V. 1985. *Vyskazyvanie i ego sootnesennost' s dejstvitel'nostju*. Moscow: Nauka

----------. 2011. “Implicitnoe otricanie y mestoimenija s otricatel’noj poliarizaciej”. *Voprosy Jazykoznanija* 1: 3–18.

----------. 2018. “Russkie mestoimenija svobodnogo vybora”. *Russian Linguistics* 42.3: 291–319.

Shvedova, Maria, von Waldenfels, Ruprecht, Yarygin, Sergey, Rysin, Andriy, Starko, Vasyl, Nikolajenko, Tymofij et al. (2017-2022): GRAC: General Regionally Annotated Corpus of Ukrainian. Electronic resource: Kyiv, Lviv, Jena. Available at https://uacorpus.org/.

Tatevosov, Sergej G. 2002. *Semantika sostavljajuščix imennoj gruppy: kvantornye slova*. Moscow: IMLI RAN.