

(C)overt (in)definiteness marking in Bulgarian, Polish, Russian, and Ukrainian

DARIA ALEXANDROVA (SERES)¹ AND LUCA MOLINARI²

¹University of Graz

²University of Wrocław & Ca' Foscari University of Venice

ABSTRACT

The paper presents an experimental investigation of NP marking of discourse-new subjects in Bulgarian, Polish, Russian, and Ukrainian. The study showed that bare NPs are favoured in articleless languages, despite the general availability of overt indefinite marking, which can be considered optional. In Bulgarian (language with a definite article) the use of bare NPs is restricted to non-salient referents (e.g., with postverbal subjects). The results obtained provide support to the hypothesis that bare NPs in articleless languages are compatible with an indefinite interpretation.

KEYWORDS · bare nominals · experimental study · indefiniteness · NP marking

1 INTRODUCTION

This paper presents an experimental investigation of nominal marking linked to the expression of indefiniteness in Slavic. The empirical data for the study come from self-reported L1 speakers of four Slavic languages: Bulgarian (South Slavic), Polish (West Slavic), Russian, and Ukrainian (East Slavic). It has traditionally been assumed that these languages lack an indefinite article (Dryer 2013, i.a.). However, there is an important difference between Bulgarian and the other three languages with respect to the presence of the article in the grammatical system. Polish, Russian, and Ukrainian have no articles at all, which means that bare nominals are expected to freely appear in argument positions in these languages. Bulgarian, on the other hand, has a definite article (in the form of an enclitic), so the use of bare NPs is expected to be restricted (although still possible, see, e.g., Gorishneva 2013).¹

One of the general questions with respect to the semantics of bare nominals in articleless languages is what interpretations are available and whether they are underlying or derived. There are several theoretical views regarding this question that should be mentioned here. Paykin & van Peteghem (2002), Dayal (2004b), among others, claim that bare NPs in languages without articles (e.g., in Russian) are underlyingly definite, which presupposes that bare singulars (although not bare plurals) in such languages cannot be *bona fide* indefinites and cannot be used in contexts that induce an indefinite interpretation. In order to achieve an indefinite interpretation, an overt marker is needed.

Another view, represented by Seres et al. (2019), Borik et al. (2020), Šimík & Demian (2020, 2021), Šimík & Burianová (2020), Seres & Borik (2021), states that bare NPs in Russian and Polish are underlyingly indefinite, and the definite interpretation becomes available only under certain conditions, very often related to pragmatics or information

¹The presence of articles and the restricted distribution of bare NPs have been claimed to be related, see the Blocking Principle elaborated by Chierchia (1998), according to which, if a language can lexicalize some operator (e.g., the iota or the existential operator), it must do so. For instance, both English and Romance have articles and still allow for bare (plural) NPs, but their distribution is strongly restricted (see, e.g., Delfitto & Schroten 1991 for some Romance languages, where bare plurals can occur in object but not in subject position - unless further restrictions apply). The case of Bulgarian is analogous: bare NPs can appear only in certain contexts (see §3.1).

structure (such as, for instance, familiarity/identifiability, ontological uniqueness, or topicality). Some researchers, for example, Chierchia (1998) and Geist (2010), among others, suggest that bare NPs in argument position are always ambiguous between a definite and an indefinite reading, and the context presumably endorses or rules out one interpretation or another. Advancing the outcome of the current experimental investigation, we can say that the empirical picture is compatible to a significant extent with either a default indefinite or an underspecified between definite and indefinite reading of bare nominals (see §5).

Another important question that we tackle here is the degree of obligatoriness/optionality of the indefiniteness marking in Slavic. This question has been raised in Geist (2013), Gorishneva (2013, 2016) for Bulgarian, Topolińska (1984), among others, for Polish, and in Padučeva (1985), Geist (2010), Borik (2016), among others, for Russian. To the best of our knowledge, there is no formal linguistic or empirical research on nominal marking in Ukrainian. The main drawback of the above-cited literature is the fact that the discussion of the obligatoriness/optionality of nominal marking is mainly based on authors' judgments and/or small groups of informants.

One of the aims of the current paper is to bring in new empirical data from several Slavic languages to contribute to the theoretical discussion of the underlying nature of bare NPs and the obligatoriness/optionality of indefiniteness marking in Slavic. In order to do this, we conducted an experimental study with self-reported L1 speakers of four Slavic languages, which is presented in the following sections. §2 is dedicated to a detailed description of our experimental investigation (research questions, design, experimental items, response annotation, etc.). In §3 we present the language-specific results obtained through the empirical study and their statistical significance. §4 contains the discussion of the results for each language and a general discussion. §5 concludes the paper.

2 THE EXPERIMENTAL INVESTIGATION

2.1 RESEARCH QUESTIONS

As stated above, the main goal of this investigation is to obtain novel empirical data in order to contribute to the theoretical discussion of bare vs marked NPs in Slavic languages. To achieve this goal, we aim to answer the following research questions:

1. What is the choice of NP form (bare vs overtly marked as indefinite or definite) for a discourse-new, not easily context-inferable singular referent in subject position in Bulgarian, Polish, Russian, and Ukrainian?
2. Is the choice of form (bare vs non-bare) affected by the following factors: type of sentence (existential/presentational vsthetic sentence), linear word order (preverbal vs postverbal subject), or modification by a relative clause?
3. Is there optionality between bare vs overtly marked NPs?
4. Do articleless languages pattern together?

It should be noted that this is an exploratory investigation, so we did not have any specific predictions with respect to the research questions. In order to answer them, we conducted an experimental investigation with self-reported L1 speakers of the four languages under study. Since there is virtually no experimental investigation on these issues (with the notable exception of Šimík & Burianová 2020, Šimík & Demian 2020, 2021), we also aim to verify the reliability of a task that has never been applied to research on such a topic, that is, the gap-filling task. We describe the experimental design in the following subsections. The answers to the research questions are provided in the general discussion (§4.5) of this paper.

2.2 DESIGN AND PARTICIPANTS

The experiment was designed as an online questionnaire consisting only of gap-filling tasks. Participants were presented with sentences with a gap before the subject NP, and they were instructed to fill the gap with any suitable expression, only if necessary. Otherwise, they could leave the gap blank (i.e., opt for the bare NP); this option required the least effort from the participants and was expected to be the easiest to choose.

In all the sentences used as experimental items, the subject NPs preceded by a gap were discourse-new, not easily context-inferable, and non-topical. This was necessary to trigger an indefinite reading of the nominal whose form was being investigated. In all the contexts, the existence of the referent was asserted, and the referent (or at least a set of potential referents) was supposed to be known to the speaker. Thus, the intended reading of the subject NP was epistemically specific, in terms of von Heusinger (2002), Ihsane (2021), that is, the speaker has a particular referent in mind.

The experiment was completely anonymous and was administered online on LimeSurvey (LimeSurvey GmbH 2012) (the interface of this platform is shown in Figure 1). 80 Bulgarian (age range: 18-68; $M=42.24$; $SD=10.95$), 89 Polish (age range: 21-74; $M=36.61$; $SD=13.16$), 85 Russian (age range: 20-68; $M=41.86$; $SD=9.06$) and 42 Ukrainian (age range: 19-65; $M=35.10$; $SD=12.48$) self-declared L1 speakers participated in the experiment.² They were recruited mainly through social networks.

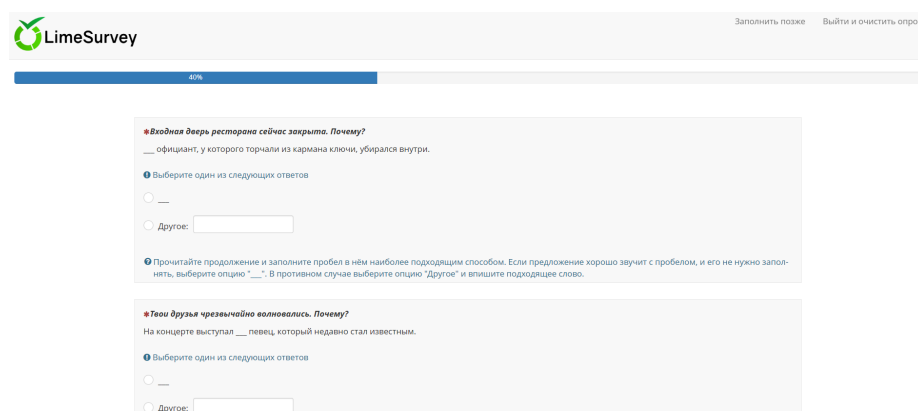


Figure 1: Lime Survey interface

Before the main task, participants filled in a short sociolinguistic questionnaire, which included the participants' age, sex, level of studies, linguistic training, and global daily use of their native language. These individual characteristics did not seem to affect the results in any significant way, meaning that the participants' answers reflect general trends in the grammar of their native language at its current stage of development.

The main experimental task consisted of 24 experimental items (4 x 6 item types, as explained in detail in §2.3) and 12 filler items (which had the same structure of experimental items but with the gap on the object NP). They were presented in randomized order for each participant, thanks to a specific function that was set in LimeSurvey before activating the questionnaire. There was a unique list of stimuli which was translated (with the help of L1 speakers) into the four languages, so that all the questionnaires contained the same experimental items and fillers.

²The number of Ukrainian participants is lower than for the other languages as it is not easy to recruit participants due to the current sociopolitical situation in Ukraine - the war. Many Ukrainian participants are war refugees, and we are immensely grateful to them for their time and effort to fill in our questionnaire.

2.3 EXPERIMENTAL ITEMS

The sentences used as experimental items, where the marking of the subject NP (bare vs non-bare) was tested, contained intransitive imperfective verbs as predicates.³ The surface morphological form of the subject NPs was nominative singular; theta-roles or other semantico-syntactic parameters were not controlled for. The sentences used as experimental items were of the following three types:

1. Existential sentences.

Sentences of this type are similar to the beginning of a fairy-tale, starting with *Once upon a time there was / there lived...* Such contexts are also called presentative, that is, a new referent is introduced into the discourse (i.e., its existence is established) and it is picked up in the subsequent discourse. The established referent is discourse-persistent as the following story is about it.

Cross-linguistically, existential sentences are a canonical environment for indefinite nominals (Carlson 1977, Dayal 2004b, i.a.) as they favor the appearance of a referent which is presumably unknown to the listener. That was the reason why such contexts were chosen to be tested experimentally.

In this study, we distinguish existential sentences fromthetic sentences par excellence (see the other two types of sentences below), as their status has been disputed in the literature. The existential sentences used as experimental items in our study represent story beginnings and, thus, convey completely new information and have no overt topic. From this regard, they can be consideredthetic in terms of Kuroda (1972). However, as Geist (2010:206) points out, such sentences contain stative (individual-level) predicates, for instance, *live*, which usually do not admit athetic interpretation (Sasse 1987, Jäger 2001, i.a.). Accordingly, story beginnings are viewed as categorical sentences containing an overt aboutness topic.

Another debatable point, with regard to existential constructions, is the syntactic role of the newly introduced NP. For the purposes of this study, we call them subjects, considering their nominative case and agreement with the verb. Nonetheless, we are aware that they are, in fact, logical and not syntactic subjects; they are also referred to as *pivots* of existential constructions (e.g., Frasson 2024, Frasson & Vaikšnoraitė 2025). However, the syntactic analysis of these constructions is outside of the scope of the present study.

2. Thetic sentences with postverbal subjects.

Postverbal subjects in articleless languages are expected to get an indefinite interpretation. The observation that the linear word order may play a role in the interpretation of subjects has been established in traditional linguistic literature on Slavic languages, for example, Russian (Pospelov 1970; Fursenko 1970; Chvany 1973, i.a.) or Czech (Krámský 1972). However, more recent studies, such as Šimík (2014) on Czech, Czardybon et al. (2014) on Polish, and Borik et al. (2020) and Seres et al. (2019) on Russian have shown that there is no clear one-to-one correlation between the syntactic position of the nominal and its interpretation. Nonetheless, we found it necessary to check for this factor (preverbal vs postverbal position of the subject) experimentally in the languages under study.

3. Thetic sentences with preverbal subjects.

Preverbal subjects in articleless languages (e.g., in Russian) are expected to get a definite interpretation, according to the traditional linguistic literature cited

³As pointed out by an anonymous reviewer, example (4) features an unergative verb, while (6) features an unaccusative verb. Since these two categories display different structures (Burzio 1986), it is possible that there is a difference between them. This is an interesting observation, as the type of intransitive was not controlled for. In designing the stimuli, we aimed at only having intransitive verbs, disregarding their type.

above and also in formal semantics (see Geist 2010 for Russian). This kind of interpretation is related cross-linguistically to the givenness/topicality of the referent (Reinhart 1981, Erteschik-Shir 2007; i.a.). In order to avoid the effect of givenness/topicality, we usedthetic sentences (thetic judgments in terms of Kuroda 1972), also known as all-focus sentences (or zero-topic sentences), in our experimental design. Notice also that the subject NPs were always discourse-new, that is the referent was novel, not given in previous discourse. Furthermore, in this study, the preverbal subject position happened to coincide with sentence-initial position, which is also relevant for NP marking.⁴

The subject in each of the three above-mentioned types of sentences was either unmodified or modified by a relative clause. It was important to control for the effect of modification as it may license certain forms and interpretations, unavailable for unmodified nominals (e.g., see Dayal 2004a for the effect of modification on the distribution of Italian bare plurals and the interpretation of English plural definites).

The combination of the three types of sentences with presence vs. absence of modification resulted in six subtypes of experimental items, which are illustrated by Ukrainian examples below.⁵

- (1) Existential sentence with a modified subject NP (EXISTENTIAL_MOD):
Žyla-bula v seli _____ babusja, jaka zbyrala korinnja i lystja. Nixto ne rozumiv, dja čogo vony byly potribni.
'In a village, there lived an old lady, who gathered roots and leaves. Nobody understood why they were needed.'
- (2) Existential sentence with an unmodified subject NP (EXISTENTIAL):
Davnym-davno žyla v zamku _____ koroleva. Vona bula nevvičlyva z usima.
'A long time ago there lived a queen in a castle. She was impolite to all.'
- (3) Thetic sentence with a postverbal modified subject NP (POSTVERBAL_MOD):
Tvoji druzi nadzvyčajno xvyľjuvalysja. Čomu? -Na koncerti vystupav _____ spivak, jakyj neščodavno stav populjarnym.
'Your friends were extremely worried. Why? -A singer who recently became famous performed in the concert.'
- (4) Thetic sentence with a postverbal unmodified subject NP (POSTVERBAL):
Tvoja kiška xvyľjuvalasja. Čomu? -Na vulyci gavkav _____ sobaka.
'Your cat was worried. Why? -In the street, a dog was barking.'
- (5) Thetic sentence with a preverbal modified subject NP (PREVERBAL_MOD):
Pered cerkvoju zbyralasja kupa ľ Judej. Ščo tam vidbuvalosja? - _____ svjaščenyk, jakyj pryjixav iz Vatykanu, propoviduvav.
'A crowd gathered in front of the church. What was happening there? -A priest who came from the Vatican was preaching.'
- (6) Thetic sentence with a preverbal unmodified subject NP (PREVERBAL):
Usi v ofisi dyvyľysja u vikno. Čomu? - _____ čolov'jaga ležav posered dorogy.
'Everybody in the office was looking outside the window. Why? -A man was lying in the middle of the road.'

⁴In a corpus study on Czech, Šimík & Burianová (2020) showed that the interpretation of a bare NP is more influenced by its (non-)sentence-initial position than its preverbal vs postverbal position.

⁵We choose to give the examples of the experimental items in Ukrainian, as, out of the four languages in our investigation, it is the least studied in the formal linguistic literature. Notice that it is not a minority language; it has around 33 million L1 speakers (Eberhard et al. 2024), but for historical and political reasons the research focus in East Slavic languages has been shifted towards Russian, while Ukrainian (as well as Belarusian) has been assumed to function in a similar way. We would like to make a small contribution to overcome this bias and promote formal approaches to Ukrainian. The translations of the same example items into Bulgarian, Polish, and Russian can be found in the Appendix.

The full list of experimental items and fillers together with the instructions given to the participants in all the languages under study can be found at <https://osf.io/p7r9e/>.

2.4 RESPONSE ANNOTATION

The answers obtained from the participants throughout the experiment were grouped and annotated under the following categories/labels:

BARE. This label was used when the participants chose to leave the gap blank. Note that this was the default option, that is, the least cognitively taxing. If participants chose to fill in the gap with a descriptive adjective (an epithet) that does not express a relation or define referential properties of the nominal (e.g., *beautiful*, *rich* etc.), the NP was still considered to be bare.

INDEF. This label was used when an overt indefiniteness marker was given as the only possible option. The indefinite markers that dominate the scene in our results can be distinguished in two categories: epistemically specific and non-specific. We define the former category as being characterized by a condition of *identifiability* in the sense of Ionin (2013:82). This amounts to saying that, whenever the speaker introduces a referent X with the use of an epistemically specific indefinite, they must be able to answer the question “Which X is it?”. This group encompasses indefinite ‘one’ in all the languages under investigation: *edin* in Bulgarian (Geist 2013, but cf. fn.10), *jeden* in Polish (Hwaszcz & Kędzierska 2018a), *odin* in Russian (Geist 2010), and *odyn* in Ukrainian. Also Polish *pewien* ‘(a) certain’ (Topolińska 1984, Kreisberg 2008) is included in this category. Epistemically non-specific indefinites, as defined here, do not obey *identifiability*, but they are rather subject to a *noteworthiness* condition (as described in Ionin 2006). This condition requires that the speaker, when using an epistemically non-specific indefinite, must have a certain degree of knowledge about the referent, without necessarily having to be able to identify/name it. This group is composed of indefinites generally translated by English ‘some’: *njakoj/njakakäv* in Bulgarian (Stankov 1995, Osenova 1999), *jakiś* in Polish (Topolińska 1984, Kreisberg 2008), *kakoj-to* in Russian (Kagan 2006, Martí & Ionin 2011), and *jakyjs’* in Ukrainian. Importantly, as mentioned in §2.2, the existence of the referent was always asserted, so all these indefinite markers are referential.

DEF. This label was used as an umbrella for markers with definiteness-related semantics, for example, demonstratives and possessives (and adjectives inflected for the definite article in Bulgarian). Such elements express a referent that is supposedly familiar or identifiable for the speaker and the hearer. The marker ‘such’ (Bulgarian *takäv*, Polish *taki*, Russian *takoj*, Ukrainian *takyi*) was also included into this group as it denotes an entity that may not be familiar but it belongs to a familiar kind of entities, that is, ‘an entity of this or that kind’.

MLTM. This label stands for “multiple-marking”. It was used when a participant indicated more than one suitable option to fill in the gap before the nominal and these options belonged to different categories (e.g., both bare and overt marking, or both definite and indefinite marking).

OTHER. This label was used for non-relevant comments that participant left (since they had absolute freedom in what they could write). However, non-relevant comments are very marginal, counting at most four instances and not in all languages.

Using this annotation scheme, we assessed the use of bare vs marked NPs for discourse-new subjects in existential andthetic sentences in the languages under study. Language-specific results are presented in the following section.

3 RESULTS

The data collected during the experimental investigation were analyzed in R (version 4.3.1). The main packages employed were *data.table* and *dplyr* (for creating tables and

calculating frequencies), ggplot2 (for plotting the graphics), and stats (for applying the Fisher's Exact Test). The following graphics report the results, indicating the percentage of occurrences for each answer type. Indicating the results as percentages (instead of giving absolute numbers) helps make the results more comparable, since the number of participants for each language was different.

3.1 RESULTS FOR BULGARIAN

Figure 2 shows the percentage of responses to the gap-filling task obtained from L1 speakers of Bulgarian. It can be seen that, for all types of experimental items, overt indefinite marking is used quite extensively, alternating with bare NPs, whose amount varies depending on the type of item.

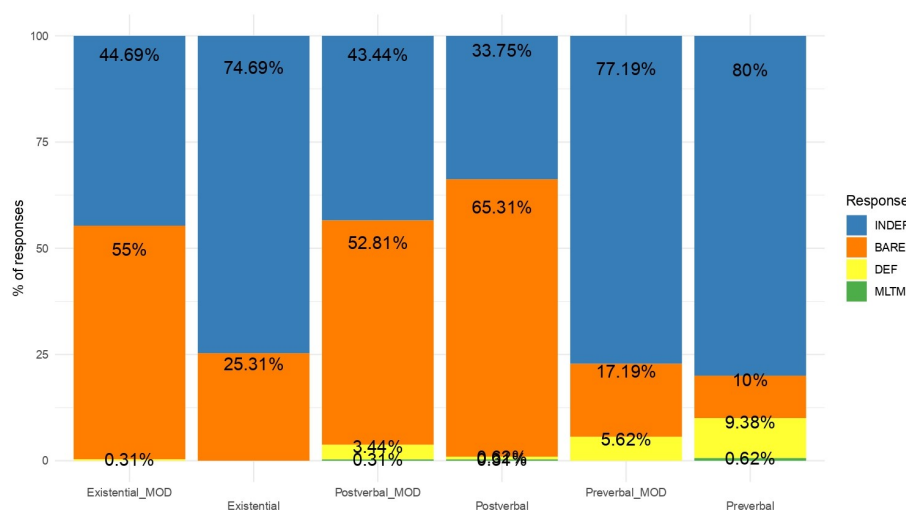


Figure 2: Results for Bulgarian

In existential sentences, modification or its absence may play a role in the choice of form of the NP. Thus, without modification, the gap before the NP is more often filled with an overt indefinite marker, while in sentences with a modified nominal, there is co-variation between bare NPs (55%) and overtly indefinite ones (44.69%).

Inthetic sentences with postverbal subjects, the appearance of overtly marked indefinites is the lowest (43.44% with modification vs 33.75% without modification). Bare NPs seem to be preferred in such contexts (52.81% with modification vs 65.31% without modification).

The highest percentage of overt indefinite marking is inthetic sentences with preverbal subjects (77.19% with modification vs 80% without modification) with bare NPs being residual (17.19% with modification vs 10% without modification).

Notice that in Bulgarian overt definite marking (by adjectives inflected for the definite article, demonstratives, or possessives) is possible, but it is marginal and is found mainly with unmodified preverbal subjects (5.62% with modification vs 9.38% without modification). Multiple (definite/indefinite) marking is very marginal (0.62% at most). Furthermore, it is interesting to notice that the unmarked form for expressing indefiniteness across-the-board is *edin* 'one'. In existential sentences, it is virtually the only form admitted. Inthetic sentences (both with post- and preverbal subject), epistemically non-specific *njakakäv* 'some' (and, to a smaller extent, *njako*j 'some') also show up.⁶

⁶The difference between *njakakäv* and *njako*j seems to be very subtle. Osenova (1999:58) exemplifies it with the example *Šte se oženja za njakoja/njakakva bogata žena* 'I will marry some(*njako*j/*njakakäv*) rich woman'. In the case of *njako*j, as Osenova observes, the speaker picks out a referent from the set of elements having the

Comparison of the distribution of different answer types was performed by means of the Fisher's Exact Test (FET).⁷ Statistically significant differences were found between the following types of items: EXISTENTIAL vs EXISTENTIAL_MOD ($p < .001$); POSTVERBAL vs POSTVERBAL_MOD ($p = .003$); PREVERBAL vs PREVERBAL_MOD ($p = .006$); POSTVERBAL_MOD vs PREVERBAL_MOD ($p < .001$); POSTVERBAL vs PREVERBAL ($p < .001$); all POSTVERBAL vs all PREVERBAL ($p < .001$).

3.2 RESULTS FOR POLISH

Figure 3 shows the results for Polish, revealing the predominant use of bare nominals in all types of experimental sentences.

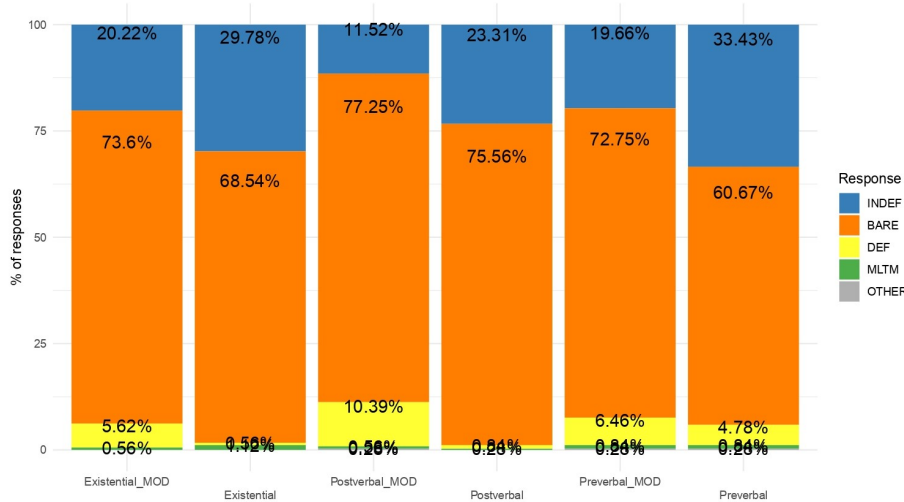


Figure 3: Results for Polish

Existential sentences show a prevalence of bare NPs (73.6% with modification vs 68.54% without modification). In such sentences, modification seems to disfavor indefinite marking (20.22% with modification vs 29.78% without modification) while boosting definite marking (5.62% with modification vs. 0.56% without modification). The indefinite marker which shows up in these presentative contexts is epistemically specific *pewien* '(a) certain'.

A similar preference for bare NPs (77.25% with modification vs 75.56% without modification) is found inthetic sentences with postverbal subjects. Here, again, modification favors definite marking (10.39% with modification vs. 0.31% without modification) at the expense of indefinite marking (11.52% with modification vs 23.31% without modification).

As far asthetic sentences with preverbal subjects are concerned, modification of the subject goes hand in hand with a larger number of bare NPs in speakers' responses (72.75% with modification vs 60.67% without modification) and definite marking (6.46% with modification vs 4.78% without modification), while indefinite marking is reduced (19.66% with modification vs 33.43% without modification). The main indefinite marker that was given by the participants inthetic sentences (both with postverbal and preverbal subject) is *jakiś* 'some', which is epistemically non-specific (Topolińska 1984, i.a.).⁸

feature "rich woman". In the case of *njakakāv*, instead, the speaker selects the set of "rich woman" from other possible sets (e.g., "poor woman") and then picks out a referent from that set.

⁷FET is more reliable than the more popular Chi-square test when the dataset contains cells with values lower than 5. This is often the case for MLTM and OTHER, which display a very low number of occurrences.

⁸As one anonymous reviewer points out, these results are surprising, as Polish *jeden* 'one' would be expected to show up, since it is regarded as more grammaticalized than 'one' in other Slavic languages (Hwaszcz &

Multiple (definite/indefinite) marking is very rare in all types of sentences (1.12% at most). This may mean that the participants interpreted the sentences as intended by the experimental design, that is, the referent of the subject NP being indefinite.

Statistically significant differences were found between EXISTENTIAL_MOD vs EXISTENTIAL (FET, $p < .001$); POSTVERBAL_MOD vs POSTVERBAL (FET, $p < .001$); PREVERBAL_MOD vs PREVERBAL (FET, $p < .001$); POSTVERBAL_MOD vs PREVERBAL_MOD (FET, $p < .001$); POSTVERBAL vs PREVERBAL (FET, $p < .001$); and all POSTVERBAL vs all PREVERBAL (FET, $p < .001$).

3.3 RESULTS FOR RUSSIAN

The responses from the L1 Russian speakers are presented in Figure 4. In Russian, the overall preferred form is bare NPs. However, there are differences regarding the sentence type.

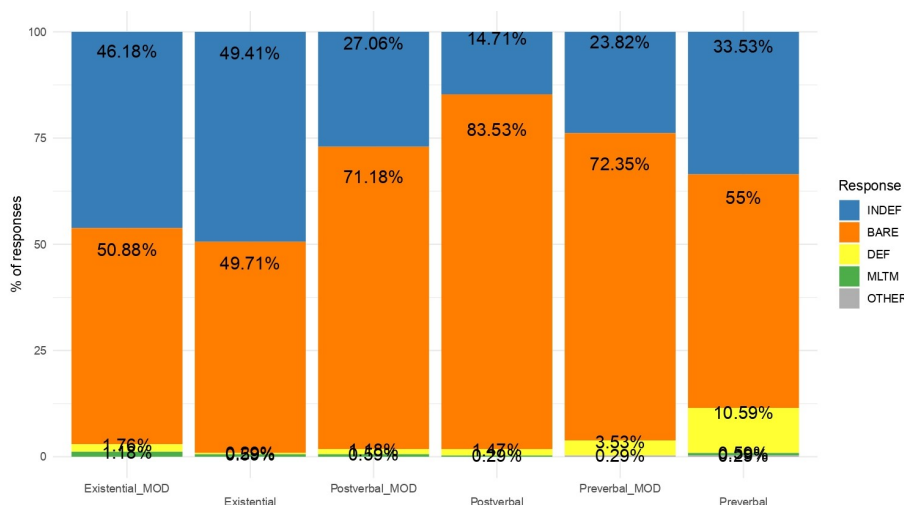


Figure 4: Results for Russian

Existential sentences show an almost equal split between bare NPs and overtly marked indefinites. Notice that the indefinite marking is almost exclusively done by means of the indefinite marker *odin* 'one'.

Inthetic sentences with postverbal subjects, the appearance of overtly marked indefinites is the lowest (27.06% with modification vs 14.71% without modification). Bare NPs are strongly preferred in such contexts (71.18% with modification vs. 83.53% without modification).

Inthetic sentences with preverbal subjects, overt indefiniteness marking is higher with unmodified referents (23.82% with modification vs. 33.53% without modification). However, bare NPs still prevail (83.53% with modification vs. 55% without modification).

A few NPs receive overt definite marking (by demonstratives or possessives). They are mainly preverbal unmodified subjects (3.53% with modification vs. 10.59% without modification). There is also a negligible amount of multiple marking (1.18% at most).

FET revealed statistically significant differences between POSTVERBAL_MOD vs POSTVERBAL ($p = .001$); PREVERBAL_MOD vs PREVERBAL ($p < .001$); POSTVERBAL vs

Kędzierska 2018a). However, we suspect that the use of *jeden* 'one' in the results of the study by Hwaszcz & Kędzierska (2018a) may be overrepresented, as *jeden* was given as a possible option to the participants. The design of our study, instead, let the relevant markers emerge. Still, in our data there are some instances of *jeden*, although they are quite marginal. Probably, *jeden* is only marginally used as an indefinite marker, at least in the written register. As noted by Hwaszcz & Kędzierska (2018b) themselves, *jeden* as an indefinite marker is likely to have specialized in the spoken register.

PREVERBAL ($p < .001$); all POSTVERBAL vs all PREVERBAL ($p < .001$). However, no significant difference is found between EXISTENTIAL_MOD vs EXISTENTIAL (FET, $p = .19$) and between POSTVERBAL_MOD vs PREVERBAL_MOD (FET, $p = .08$).

3.4 RESULTS FOR UKRAINIAN

Figure 5 shows that in Ukrainian, as well as in Polish, the most commonly used form in all experimental contexts is a bare NP. The use of bare NPs is more prominent than in Russian.

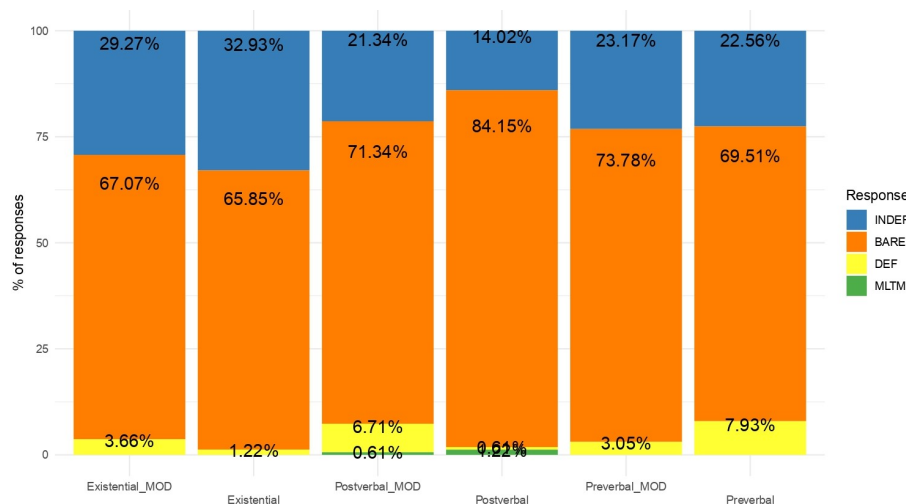


Figure 5: Results for Ukrainian

As far as existential sentences are concerned, there is a predominance of bare NPs irrespective of the presence of modification (66.07% with modification vs 65.85% without modification). However, the percentage of overt indefinite marking is still not negligible (29.27% with modification vs. 32.93% without modification). The favored form to mark indefiniteness in such contexts is *odyn* 'one'.

Inthetic sentences with postverbal subjects, the appearance of overtly marked indefinites is the lowest (21.34% with modification vs. 14.07% without modification). Bare NPs are preferred in such contexts (71.34% with modification vs. 84.15% without modification).

Inthetic sentences with preverbal subjects, there is still a preference for bare NPs (73.78% with modification vs. 69.51% without modification). Overt indefinite marking is, however, not marginal (23.17% with modification vs. 22.56% without modification).

Overt definite marking (by demonstratives or possessives) is generally marginal, with a peak of 7.93% in preverbal position with unmodified subjects (as was the case for Bulgarian and Russian). Multiple marking is found in postverbal position, although it is very residual (with a peak of 1.22% in PREVERBAL).

The statistically significant differences detected by FET emerged between POSTVERBAL_MOD vs POSTVERBAL ($p = .002$) and between POSTVERBAL vs PREVERBAL ($p < .001$). No significant difference was found in the following pairs: EXISTENTIAL_MOD vs EXISTENTIAL (FET, $p = .32$), PREVERBAL_MOD vs PREVERBAL (FET, $p = .19$), POSTVERBAL_MOD vs PREVERBAL_MOD ($p = .32$), and all POSTVERBAL vs all PREVERBAL ($p = .068$).

4 DISCUSSION

In this section, we first discuss language-specific results and then move on to discussing the overall outcomes and answering the research questions of this experimental investi-

gation.

4.1 BULGARIAN

According to the results obtained in the experimental study, linear syntactic position turns out to be significant for overt indefinite marking in Bulgarian. Preverbal subjects are predominantly marked, while postverbal subjects are instead mainly bare. This may happen because preverbal subjects favor a higher degree of discourse salience and identifiability, as suggested in Ionin (2013) for Russian, and therefore do not require any reference-establishing marking.

A rather high level of overt indefiniteness marking in existential sentences can be explained by the presence of discourse anaphora (and thus easier identifiability of the referent) in such contexts: a referent is introduced by an existential sentence and is picked up in the following sentence. However, modified subjects of existential sentences had a lower overt marking, compared to the sentences with non-modified subjects. This may be due to the absence of the need to identify the referent by an overt indefinite, as it is identified through modification. It is not clear, though, why this effect was only found in existential sentences.

It is important to notice that the only indefinite marker available in existential sentences is *edin* 'one'. 'One'-marked indefinites in Bulgarian have been claimed to denote specific referents (Gorishneva 2013, 2016), which is the case of subjects of existential sentences and their discourse persistence.⁹

Inthetic sentences, 'one' still dominates the scene, but a discrete amount of *njakakäv* 'some' is also found. The use of 'some' can be associated with epistemic non-specificity, that is, the speaker is aware of the existence of the referent but is not aware of its identity.¹⁰

Bare NPs in Bulgarian have been claimed to denote non-specific/non-identifiable referents, which is compatible with their presence in postverbal position, associated with lower discourse prominence (see Gorishneva 2013, 2016 for a characterization of bare NPs vs *edin*-NPs in Bulgarian in terms of type- vs token-reference, respectively).

In Bulgarian, the competition between bare vs overtly marked NPs is limited to the indefiniteness domain alone. In fact, as Bulgarian has a definite article, the definite interpretation of a bare NP is blocked (see Chierchia 1998's Blocking Principle; see also fn.1).

4.2 POLISH

As was shown in §3.2, in Polish, bare NPs are the preferred form in all the experimental contexts.

⁹An anonymous reviewer asks whether we considered genericity for existentials. We however contend that existentials are not generics, which are traditionally analyzed as law-like statements (see Seres 2020 for an overview on genericity). The reviewer points out that in Serbian, an existential sentence like (i) only admits the indefinite *neki* 'some', not *jedan* 'one'. They link this to the fact that the students in 9 are part of the discourse universe (so they are not indefinite *stricto sensu*), yet not uniquely identified. In Bulgarian, 'one' is allowed in a context such as (ii).

- | | | |
|------|---|-----------|
| (i) | Tamo ima *jednih / nekih studenata
there have.3SG *one.GEN.PL / some.GEN.PL student.GEN.PL
'There are some students there.' | Serbian |
| (ii) | Tam ima edni studenti.
there have.3SG one.PL student.PL
'There are some students there.' | Bulgarian |

¹⁰Geist (2013) argues that Bulgarian *edin* 'one', as an indefinite, must obey the condition of *identifiability* (in the sense of Ionin 2013:82; see §2.4). However, the presence of *edin* 'one' as the unmarked indefinite form also inthetic sentences goes against this claim, as the referent is generally discourse-new and unfamiliar for the speaker. This lends support to the observation made by Molinari (2023) that the behavior of *edin* would be better characterized in terms of *noteworthiness* (in the sense of Ionin 2006; see §2.4).

It is important to note that, despite the significant differences, linear syntactic position does not dramatically change the distribution of NP marking. This is in line with experimental findings by Šimík & Demian (2021), who showed that pre-/postverbal position does not appear to have any effect on the interpretation of the NP subject. This result can also be compared to the results obtained for Russian bare plural subjects by Seres et al. (2019), who showed experimentally that there is no one-to-one correspondence between the linear syntactic position of a bare NP and its interpretation, but information structure factors and discourse prominence may come into play.

Modification in its turn seems to have a uniform influence on marking in Polish, favoring both bare and definite NPs at the expense of indefinite ones. This may be due to the identification function that the modification performs, which enhances the degree of familiarity of the referent for the speaker.

There is also a visible effect of sentence type regarding the choice of indefinite markers. In fact, in existential sentences, only the epistemically specific marker *pewien* '(a) certain' is admitted. Inthetic sentences (with both post- and preverbal subject), on the other hand, the main marker given by the participants is epistemically non-specific *jakiś* 'some'. This can be related to higher discourse prominence and discourse persistence of the referent in existential sentences as compared tothetic sentences.

4.3 RUSSIAN

In Russian, bare NPs are the preferred form for subjects in indefiniteness-inducing contexts, regardless of the sentence type and the linear syntactic position of the subject. However, the overall amount of overt indefinite marking is not negligible, which may mean that there is real optionality in the use of (non-)bare NPs. The relatively low use of overt indefinites (as compared to bare NPs) also indicates that indefiniteness marking cannot be obligatory in indefiniteness-inducing contexts (contra Dayal 2004b, 2011). The results of the current study corroborate theoretical predictions regarding the non-obligatoriness of indefinite marking for indefinite NPs in Russian (e.g., by Borik 2016) as well as the experimental results of Seres et al. (2023).

There is an effect of sentence type in the choice of indefinite marking: different markers are used to introduce subject NPs in existential presentative sentences and inthetic sentences. In existential sentences, the only marker that shows up is *odin* 'one', which is known to be epistemically specific (Padučeva 1985, Geist 2008, 2010, Pereltsvaig 2007, Ionin 2013, Borik 2016). The use of overtly marked indefinites was highest in existential sentences, which may be due to discourse prominence of the referent (the referent was picked up in the following discourse); thus, it was rather easy to identify. Licensing of *odin* in existential sentences can be explained by the identifiability condition (Ionin 2013): *odin* as a specificity marker can only appear on NPs in contexts where the referent can be identified by the speaker, that is, the referent is epistemically specific (in terms of von Heusinger 2002; Ihsane 2021).

Inthetic sentences (with post- and preverbal subject), instead, the epistemically non-specific marker *kakoj-to* 'some' (Kagan 2006, 2011, Martí & Ionin 2011) appears along with *odin*. The two markers appear at roughly the same rate. Interestingly, however, in postverbal subject position, *odin* appears in the contexts with modified subject, while *kakoj-to* is concentrated in the contexts without modification (this tendency does not clearly emerge in preverbal position). Such a distribution of indefiniteness marking resembles the one in Polish and seems to correspond to epistemic specificity/non-specificity of the referent. Furthermore, the appearance of *odin* with modified subjects naturally follows from its specificity, as the subject can easily be interpreted as identifiable, since its possible reference is already restricted by the relative clause. In general, however, the use of overt indefinites was lowest in postverbal unmodified subjects ofthetic sentences, that is, in contexts where the speaker may not be able to identify the referent (even though they are aware of its existence).

As for definiteness marking, we observe a boost in preverbal position. This goes in the direction of the observation Geist (2010) made for Russian, that preverbal subjects are generally interpreted as definite (as the preverbal position generally coincides with topic). Despite this, the amount of definite marking is too low to provide strong support to Geist's theory.

It can be suggested that, in general, overt indefinite (and also definite) marking appears on the nominal in Russian when disambiguation is needed for pragmatic reasons. Otherwise, a bare nominal is used.

4.4 UKRAINIAN

In Ukrainian, bare NPs are the preferred form in all the experimental contexts, like in Russian and Polish. However, considering the total percentage of bare and of overt indefinite marking, Ukrainian is closer to Polish than to Russian. The total percentage of bare NPs in Ukrainian is 71.95%, similar to the total 71.40% registered in Polish. Russian displays a lower percentage of bare NPs, around 63.77%. As for overt indefinite marking, Russian features a higher total percentage (32.45%) than the other two articleless languages, which instead pattern together (22.99% in Polish, and 23.88% in Ukrainian).

Just like in Russian, linear syntactic position in Ukrainian seems to affect the distribution of NP marking. Namely, in preverbal position, there is a larger amount of overt definite marking.

Modification seems to uniformly promote (although very slightly) overt indefinite marking, but only inthetic sentences. This tendency is opposite to what happens in Polish. This may be due to the identification function that the modification performs.¹¹ Modification seems to smooth out the differences between the post- and preverbal position, which display no statistically significant difference (which was instead found between unmodified postverbal vs preverbal subjects).

Similarly to what happens with the other languages, there is a visible effect of sentence type on the choice of indefinite markers. Existential sentences mainly feature marking by *odyn* 'one',¹² while inthetic sentences the epistemically non-specific marker *jakyjs* 'some' dominates the scene. These results are not surprising. Furthermore, their reliability is confirmed by the fact that they replicate the state of affairs in the other three languages.

4.5 GENERAL DISCUSSION

After discussing language-specific results, here we turn to answering the research questions stated in §2.1 based on the data from our experimental study. The research questions are repeated below for convenience:

1. What is the choice of the NP form (bare vs overtly marked as definite or indefinite) for a discourse-new referent in subject position in Bulgarian, Polish, Russian, and Ukrainian?

Speakers of articleless languages (Polish, Russian, and Ukrainian in our sample) prefer bare NPs in all the contexts investigated, while Bulgarian speakers favor them only in unmodified existential sentences and in postverbal position. In other cases, Bulgarian displays overt indefinite marking as the default option.

2. Is the choice of form (bare vs non-bare) affected by the following factors: type of sentence (existential vs non-existential), linear word order (preverbal vs postverbal subject), or modification by a relative clause?

¹¹We are aware that some effects that are found are likely to be linked to the relatively small dataset. Verification of the effect of modification is left for future research.

¹²This result is particularly relevant, as it provides further support to the conclusion reached by Seres (2024) that the numeral 'one' in Ukrainian is undergoing a process of grammaticalization, *contra* Heine & Kuteva (2006) who argue that *odyn* 'one' can only be used as a numeral.

Overall, (i) type of sentence, (ii) linear order, and (iii) modification seem to influence NP marking in different ways:

(i) The type of sentence seems to influence the overall marking only in Bulgarian (as modification has a strong effect in reducing indefinite marking only in existential sentences) and Russian (as bare NPs and overt indefinite marking are present to the same extent only in existential sentences). This condition, however, affects the choice of indefinite marker in all the languages. Existential sentences favor epistemically specific markers, that is, Bulgarian *edin* 'one', Polish *pewien* '(a) certain', Russian *odin* 'one', Ukrainian *odyn* 'one'. Inthetic sentences, epistemically non-specific markers show up instead, that is, Bulgarian *njakakäv* 'some' (together with *edin*), Polish *jakiś* 'some', Russian *kakoj-to* 'some' (together with *odin*), Ukrainian *jakyjs* 'some'.

(ii) In general, NPs in preverbal position get more (in)definite marking than those in postverbal position, which in all four languages are generally left bare.¹³ Bare postverbal subjects may be linked to the perceived reduced discourse prominence of the relevant NP. In contrast, perceived discourse prominence is boosted in preverbal position. On the one hand, the higher percentage of indefinite marking in preverbal position in the three articleless languages goes in the direction of Geist (2010)'s prediction (worked out for Russian) that preverbal bare singulars are by default interpreted as definite. For this reason, overt indefinite marking is needed to trigger an indefinite interpretation. Moreover, Geist claims that modification of a bare preverbal singular NP makes an indefinite reading available. Indeed, in Polish and Russian, preverbal modified subjects received less overt indefinite marking than unmodified ones.¹⁴ On the other hand, the large amount of bare NPs in preverbal position does not allow us to firmly support Geist's theory. There exists a possibility that the NPs that were left bare in preverbal position were actually interpreted as being definite. Our data do not allow us to draw such a conclusion, so more research will be needed to experimentally verify this prediction.

(iii) No clear role of modification emerges from the data presented. In languages like Bulgarian and Russian, modification seems to enhance indefinite marking in postverbal position and to discourage it in preverbal one. In Polish, modification seems to consistently reduce indefinite marking, whereas in Ukrainian it has the opposite effect (inthetic sentences). More research is needed on this issue.

3. Is there optionality between bare vs. overtly marked NPs?

The answer to this question hinges on the definition of "optionality" one assumes. If one takes optionality to indicate equal chances of obtaining either a bare or an overtly marked NP, then the only true case of optionality is instantiated by existential sentences in Russian. However, there is another understanding of optionality. As pointed out by Sorace (2000: 94), optionality "involves the coexistence of two (or more) optimal forms, one of which is usually 'more grammatical' than the other". This definition of optionality is more concerned about the existence of multiple forms for expressing the same meaning (rather than the probability of obtaining either), and also leaves room for the existence of a preference for one form over the other(s). In this sense, one could then observe that optionality in NP marking is generalized. In articleless languages, overt marking is always available

¹³An anonymous reviewer wonders about the role of theta roles and case in these contexts. As mentioned in 2.3, thematic roles of the subjects were not controlled for (as they are tied to the unaccusativity/unergativity of the verb). As for case, in all the sentences, the subject shows up in the nominative. Since the distribution of marking (and of the kind of markers) seems to be mainly influenced by context type and subject position, we have no evidence of an active role of thematic roles (although a dedicated study would be needed to verify this).

¹⁴Recall that FET did not detect any significant difference between PREVERBAL_MOD vs PREVERBAL in Ukrainian. More research is needed to verify whether the same tendency is found in this language as well.

in all contexts. In Bulgarian, the only article language in our sample, bare NPs are often available (with the exception of subjects in preverbal position). This, however, does not exclude that articleless languages prefer bare NPs over overtly marked ones (and the same holds for Bulgarian, mainly with postverbal subjects).

There can be yet another, purely formal, understanding of syntactic optionality, as described in Biberauer & Richards (2006). In their terms, ‘pure optionality’ (i.e., alternative structures that do not differ semantically) arises as a consequence of the existence of multiple equally costly derivations that satisfy an EPP feature.¹⁵ An in-depth formal discussion of the application of this proposal to the case at hand would fall beyond the scope of this paper. Let us just briefly mention a possible (very tentative) way to reconcile our data with Biberauer and Richard’s theory. Following Longobardi (1994), the DP-layer needs to be filled to make the noun referential. This could be achieved by either merging some element in that layer, or (c)overtly moving the noun there. Giusti & Iovino (2016) argue that in Latin, N could procrastinate the movement to D because rich case morphology made the N-to-D link visible. In the same spirit, one could argue that the rich case morphology in Polish, Russian, and Ukrainian makes it possible to procrastinate N-to-D movement at LF. This option competes with the merger of an indefinite marker in the DP-layer in overt syntax. LF-movement being considered more economical (Chomsky 1995), this operation can be regarded as being as costly as the overt merger of an element in the DP-layer. As for Bulgarian postverbal bare NPs, this option may be due to the fact that this position is governed (Chomsky 1981). This issue is, however, left open for future research.

4. Do articleless languages pattern together?

Articleless languages do pattern together if one considers the general preference for bare NPs (with the exception of existential sentences in Russian, which could be due to a crystallized formula). However, a distinction can still be made inside this group. In fact, Polish and Ukrainian seem to be closer to each other than Russian. Russian ranks highest in the use of overt indefinite marking (with a total of 32.45%), with Polish and Ukrainian displaying a lower value (22.99% and 23.88%, respectively). This order is reversed if one considers bare NPs, with Ukrainian and Polish featuring the highest percentage (71.40% and 71.95%, respectively) which decreases in Russian (63.77%). Interestingly, the postverbal position is the most homogeneous one, as it is characterized by a prevalence of bare NPs in all the languages, including Bulgarian.

In general, it is important to note that a definite reading of a bare NP is never excluded in articleless languages. Despite this, indefiniteness-inducing contexts seem to have effectively triggered an indefinite reading of the NP subject. This is proven by the low presence of definite marking in Bulgarian. In this case, Bulgarian functions as a touchstone for the interpretation of the subject NPs in our experimental items. In fact, as mentioned in §4.1, a definite reading is only achievable by overt definite marking in Bulgarian. As a consequence, the low percentage of definite marking signals that the referents were almost exclusively interpreted as being indefinite.

It is also important to point out that our findings are in line with the results of the recent experimental study by Seres et al. (2023), who investigated the perceived naturalness of bare/*odin* NPs in Russian and Catalan NPs introduced by the definite/indefinite article in bridging contexts. Their experiment presented contexts which introduced either one or two referents. Seres et al.’s investigation supported

¹⁵For concrete examples, the reader is directed to the paper itself. We thank an anonymous reviewer for suggesting that we consider this view of optionality as well.

the view that bare NPs in Russian are compatible with both a definite and an indefinite reading, as their perceived naturalness was comparable to that of Catalan NPs introduced by either the definite or the indefinite article. Furthermore, their study confirmed the status of *odin* as a specific marker for the introduction of new referents. Despite this, their results showed that bare NPs are also suitable for introducing discourse-new referents, going in the direction of non-obligatoriness of *odin* (Pađučeva 1985, Borik 2016). Our experimental investigation provides further support for these conclusions, which are arrived at through a different experimental task. This also proves the validity of gap-filling for investigating NP marking.

5 CONCLUSIONS

This study has shown that the gap-filling task is a quite reliable and powerful tool to investigate the distribution of (non-)bare NPs. The reliability of the data lies in the fact that the “bare” option was presented as a default, that is, overt marking was deemed necessary by the subjects who filled in the blank. Furthermore, the fact that the results of our study are in line with the conclusions of previous investigations provides evidence in favor of the validity of such a task.

As this study showed, overt indefinite marking is available (to different extents) to articleless languages in all contexts. This indicates the presence of optionality, where one form (in this case, bare NPs) is preferred over the “competing” one (i.e., overt indefinite marking). This hints at the fact that bare NPs may not be “bare”, but rather headed by a null indefinite (which competes with the overt ones; see Borik & Espinal 2019 for the idea of a null D heading bare singular subjects of kind-level predicates in Russian).

The sporadic use of definiteness- or multiple marking on NPs in all languages may indicate that indefiniteness and non-topicality induced by the context can be overridden by the participants’ interpretation of the situation and the salience of the referent in it.

Although the languages lacking articles in our sample generally prefer bare NPs, there are still certain differences in the use of overt (in)definite marking. In particular, Russian displays the highest percentage of overt indefinite marking, followed by Polish, with Ukrainian being ranked last as it displays the least amount of overt indefinite marking. For definite marking, however, the ranking is reversed, with Ukrainian sitting in the first place for overt definite marking, followed by Polish, and Russian in the third place. This discrepancy in the use of (in)definite marking is worth further investigating, and it is left for future research.

Overall, the empirical data obtained support the hypothesis that bare NPs are inherently indefinite (Šimík & Demian 2020, Seres & Borik 2021, i.a.), or are at least compatible with the view that bare NPs are underspecified for (in)definiteness (Chierchia 1998; Geist 2010, i.a.). In fact, if bare NPs were underlyingly definite (as suggested by Paykin & van Peteghem 2002, Dayal 2004b), we would expect overt indefinite marking to be the unmarked form in all contexts. Recall that this conclusion holds for articleless languages only, as Bulgarian bare NPs cannot be given a definite interpretation (given the availability of a definite article).

Another relevant aspect of the present study, beyond the confirmation of previous observations in the literature, lies in the fact that the same conclusions reached for the better-studied Polish and Russian were shown to hold for Ukrainian, on which there has been virtually no investigation, as well. The present study will hopefully open the way to further research on mostly-overlooked Slavic languages, such as Ukrainian and Belarusian.

ABBREVIATIONS

3	third person	PL	plural
FET	Fisher's Exact Test	SD	standard deviation
GEN	genitive	SG	singular
M	Mean		

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for their useful comments that helped to improve our paper, and the audiences of the workshop “Issues in the Syntax and Semantics of Definiteness” (University of Bucharest, 8-9/12/2023), LE2024 (University of Potsdam, 22-23/02/2024), FASL33 (Dalhousie University & Saint Mary's University, 16-19/05/2024), and the workshop “Functional categories, dimensions of meaning, and expletiveness” (Universitat Autònoma de Barcelona, 12-14/06/2024) for their questions and suggestions regarding this study. We are grateful to Ainur Kakimova, Beata Milewicz, Karolina Zuchewicz, Ljudmila Geist, Mariya Donska, Oleksandra Protsenko, Olga Borik, Paweł Rutkowski, Tihomir Rangelov, and Vesela Simeonova for their help in translating (and commenting on) the materials for the experiments, and for helping us to recruit participants for the study. We deeply appreciate the time and effort of every native speaker who took part in the study. The present study and the required forms received ethical approval from the Research Ethics Committee at Ca' Foscari on 05.02.2020, protocol n. 1/2020.

Daria Alexandrova (Seres) acknowledges the financial support of the projects PID2020-112801GB-I00 and PID2023-150347NB-I00 of the Spanish MICIU. Luca Molinari acknowledges current financing by the OPUS 26 grant (2023/51/B/HS2/02217) from the Polish National Science Center (NCN).

CONTACT

DARIA ALEXANDROVA (SERES) — dariaseres@gmail.com
 LUCA MOLINARI — luca.molinari@uwr.edu.pl

REFERENCES

- Biberauer, Theresa & Marc Richards. 2006. True optionality: When the grammar doesn't mind. In Cedric Boeckx (ed.), *Minimalist essays*, 35–67. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.91.08bib>.
- Borik, Olga. 2016. Constraints on the position and interpretation of bare singular indefinites in Russian. *Linguistica* 56(1). 9–23. <https://doi.org/10.4312/linguistica.56.1.9-23>.
- Borik, Olga, Joan Borràs-Comes & Daria Seres. 2020. Preverbal (in)definites in Russian: An experimental study. In Kata Balogh, Anja Latrouite & Robert D. Van Valin, Jr. (eds.), *Nominal anchoring: Specificity, definiteness, and article systems*, 51–80. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.4049681>.
- Borik, Olga & Maria Teresa Espinal. 2019. Definiteness in Russian bare nominal kinds. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*, 293–318. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.3252023>.
- Burzio, Luigi. 1986. *Italian syntax: A government-binding approach*. Dordrecht: Springer Science & Business Media. <https://doi.org/10.1007/978-94-009-4522-7>.

- Carlson, Gregory N. 1977. *Reference to kinds in English*. Amherst, MA: University of Massachusetts, Amherst dissertation. <https://scholarworks.umass.edu/dissertations/AAI7726414>.
- Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural language semantics* 6(4). 339–405. <https://doi.org/10.1023/a:1008324218506>.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chvany, Catherine V. 1973. On the role of presuppositions in Russian existential sentences. In Claudia Corum & T. Cedric Smith-Stark (eds.), *Papers from the 9th regional meeting of the Chicago Linguistic Society*, 68–77. Chicago, IL: Chicago Linguistic Circle.
- Czardybon, Adrian, Oliver Hellwig & Wiebke Petersen. 2014. Statistical analysis of the interaction between word order and definiteness in Polish. In Adam Przepiórkowski & Maciej Ogrodniczuk (eds.), *Advances in natural language processing: NLP 2014*, 144–150. Cham: Springer. https://doi.org/10.1007/978-3-319-10888-9_15.
- Dayal, Veneeta. 2004a. Licensing by modification. *Ilha do desterro* 47. 217–238.
- Dayal, Veneeta. 2004b. Number marking and (in)definiteness in kind terms. *Linguistics and philosophy* 27. 393–450. <https://doi.org/10.1023/B:LING.0000024420.80324.67>.
- Dayal, Veneeta. 2011. Bare noun phrases. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of linguistics and communication science*, 1088–1108. Berlin: De Gruyter. <https://doi.org/10.1515/9783110255072.1088>.
- Delfitto, Denis & Jan Schroten. 1991. Bare plurals and the number affix in DP. *Probus* 3(2). 155–185.
- Dryer, Matthew S. 2013. Indefinite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/38>.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2024. *Ethnologue: Languages of the world*. Dallas, TX: SIL International 27th edn. <http://www.ethnologue.com>.
- Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press. <https://b-ok.cc/book/949201/752c5d>.
- Frasson, Alberto. 2024. Agreement, case and definiteness in Balkan existential constructions. *Balkanica et slavica* 4(1). 13–34. <https://doi.org/10.30687/BES/2785-3187/2024/01/001>.
- Frasson, Alberto & Elena Vaikšnoraite. 2025. There is more about existentials: Agreement and case assignment in Serbian and Lithuanian. *Journal of Slavic linguistics* 33(FASL issue). Proceedings of FASL 32, edited by Wayles Browne, George Fowler, and Catherine Rudin.
- Fursenko, Diana I. 1970. Porjadok slov kak odno iz sredstv vyraženiya neopredelennosti/opredelennosti imën suščestvitel'nyx. *Russkij jazyk za rubežom* 4. 68–72.
- Geist, Ljudmila. 2008. Specificity as referential anchoring: Evidence from Russian. *Proceedings of Sinn und Bedeutung* 12. 151–164. <https://doi.org/10.18148/sub/2008.v12io.582>.

- Geist, Ljudmila. 2010. Bare singular NPs in argument positions: Restrictions on indefiniteness. *International review of pragmatics* 2(2). 191–227. <https://doi.org/10.1163/187731010X528340>.
- Geist, Ljudmila. 2013. Bulgarian ‘edin’: The rise of an indefinite article. In Uwe Junghanns, Dorothee Fehrmann, Denisa Lenertová & Hagen Pitsch (eds.), *Formal Description of Slavic Languages: The ninth conference. Proceedings of FDSL 9, Göttingen 2011*, 125–148. Frankfurt am Main: Peter Lang.
- Giusti, Giuliana & Rossella Iovino. 2016. Latin as a split-DP language. *Studia linguistica* 70(3). 221–249. <https://doi.org/10.1111/stul.12045>.
- Gorishneva, Elena. 2013. Bare vs. non-bare nouns. Two kinds of indefinites in Bulgarian. In Johannes Kabatek & Albert Wall (eds.), *New perspectives on bare noun phrases in Romance and beyond*, 301–328. Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.141>.
- Gorishneva, Elena. 2016. *The variety of functions of the numeral and indefinite marker “one” in Bulgarian and Russian*. Herne: Gabriele Schäfer Verlag.
- Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe*. Oxford/New York: Oxford University Press.
- von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of semantics* 19(3). 245–274. <https://doi.org/10.1093/jos/19.3.245>.
- Hwaszcz, Krzysztof & Hanna Kędzierska. 2018a. The rise of an indefinite article in Polish: An appraisal of its grammaticalisation stage (Part 1). *Studies in Polish linguistics* 13. 93–121. <https://doi.org/10.4467/23005920SPL.18.005.8744>.
- Hwaszcz, Krzysztof & Hanna Kędzierska. 2018b. The rise of an indefinite article in Polish: An appraisal of its grammaticalisation stage (Part 2). *Studies in Polish linguistics* 13. 145–166. <https://doi.org/10.4467/23005920SPL.18.005.8744>.
- Ihsane, Tabea. 2021. Telicity, specificity, and complements with a “partitive article” in French. In Tabea Ihsane (ed.), *Disentangling bare nouns and nominals introduced by a partitive article*, 227–261. Leiden: Brill. https://doi.org/10.1163/9789004437500_008.
- Inonin, Tania. 2006. This is definitely specific: Specificity and definiteness in article systems. *Natural language semantics* 14(2). 175–234. <https://doi.org/10.1007/s11050-005-5255-9>.
- Inonin, Tania. 2013. Pragmatic variation among specificity markers. In Cornelia Ebert & Stefan Hinterwimmer (eds.), *Different kinds of specificity across languages*, 75–103. Dordrecht: Springer.
- Jäger, Gerhard. 2001. Topic-comment structure and the contrast between stage level and individual level predicates. *Journal of semantics* 18(2). 83–126. <https://doi.org/10.1093/jos/18.2.83>.
- Kagan, Olga. 2006. Specificity as speaker identifiability. In Beáta Gyuris, László Kálmán, Chris Piñón & Károly Varasdi (eds.), *Proceedings of the Ninth Symposium on Logic and Language*, 82–89. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences.
- Kagan, Olga. 2011. On speaker identifiability. *Journal of Slavic linguistics* 19(1). 47–84. <https://doi.org/10.1353/jsl.2011.0008>.
- Kreisberg, Alina. 2008. Ancora a proposito dei “prenomi” polacchi. *Linguistica e filologia* 26. 31–55. https://doi.org/10.6092/LeF_26_p31.

- Krámský, Jiří. 1972. *The article and the concept of definiteness in language*. The Hague: Mouton. <https://doi.org/10.1515/9783110886900>.
- Kuroda, Sige-Yuki. 1972. The categorical and the thetic judgment: Evidence from Japanese syntax. *Foundations of language* 9(2). 153–185. <https://www.jstor.org/stable/25000656>.
- LimeSurvey GmbH. 2012. *LimeSurvey: An Open Source survey tool*. LimeSurvey Project Hamburg, Germany. <https://www.limesurvey.org>.
- Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic inquiry* 25(4). 609–665. <https://www.jstor.org/stable/4178880>.
- Martí, Luisa & Tania Ionin. 2011. Wide scope indefinites in Russian: an experimental investigation. *Glossa* 4(1). 1–31. <https://doi.org/10.5334/gjgl.612>.
- Molinari, Luca. 2023. The syntax of Bulgarian edin ‘one’. *Journal of Slavic linguistics* 31(1–2). 161–213. <https://doi.org/10.1353/jsl.2023.a936022>.
- Osenova, Petya. 1999. *Semantika i funkcionirane na neopredelitelnite mestoimenija v balgarskia ezik*. Sofia: University of Sofia dissertation. <https://bultreebank.org/wp-content/uploads/2017/05/DisertaciaPetya.pdf>.
- Padučeva, Elena V. 1985. *Vyskazyvanie i ego sootnesionnost’ s deistvitel’nost’u: Referentsial’nye aspekty semantiki mestoimenii*. Moscow: Nauka.
- Paykin, Katia & Marleen van Peteghem. 2002. Definiteness in a language without articles: A case study of Russian. *Recherches linguistiques de Vincennes* 31. 97–112.
- Pereltsvaig, Asya. 2007. The universality of DP: A view from Russian. *Studia linguistica* 61(1). 59–94. <https://doi.org/10.1111/j.1467-9582.2007.00129.x>.
- Pospelov, Nikolaj. 1970. O sintaksičeskom vyražanii kategorii opredelënnosti – neopredelënnosti v sovremennom russkom jazyke. In Timofej Lomtev & Aleksandra Kamynina (eds.), *Issledovanija po sovremennomu russkomu jazyku*, 182–189. Moskva: Izdatel’stvo Moskovskogo universiteta.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94. <https://www.philosophica.ugent.be/wp-content/uploads/full-texts/27-4.pdf>.
- Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25(3). 511–580. <https://doi.org/10.1515/ling.1987.25.3.511>.
- Seres, Daria. 2020. *The expression of genericity in languages with and without articles. Russian as compared to English and Romance*. Barcelona: Universitat Autònoma de Barcelona dissertation. <http://hdl.handle.net/10803/670853>.
- Seres, Daria. 2024. The grammaticalization of numeral ‘one’ in Slavic: from quantification to (non-)referentiality. *Catalan journal of linguistics* 23(Special issue: Semantic change in grammaticalization). 1–16. <https://doi.org/10.5565/rev/catjl.443>.
- Seres, Daria & Olga Borik. 2021. Definiteness in the absence of uniqueness: the case of Russian. In Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns & Hagen Pitsch (eds.), *Advances in formal Slavic linguistics 2018*, 339–363. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.5155544>.
- Seres, Daria, Joan Borràs-Comes & Olga Borik. 2019. Interplay between position and interpretation: An experimental study of Russian bare plurals. *Revue roumaine de linguistique* 64(2). 163–177.

- Seres, Daria, Joan Borràs-Comes & M. Teresa Espinal. 2023. Bridging inferences and reference management: Evidence from an experimental investigation in Catalan and Russian. *Language and speech* 67(1). 203–227. <https://doi.org/10.1177/00238309231173337>.
- Šimík, Radek. 2014. Definiteness and articleless languages. Handout, Letní škola lingvistiky, Dačice.
- Šimík, Radek & Markéta Burianová. 2020. Definiteness of bare NPs as a function of clausal position: A corpus study of Czech. In Tania Ionin & Jonathan MacDonald (eds.), *Proceedings of Formal Approaches to Slavic Linguistics (FASL) 26: The Urbana-Champaign meeting 2017*, 343–361. Ann Arbor, MI: Michigan Slavic Publications.
- Šimík, Radek & Christoph Demian. 2020. Definiteness, uniqueness, and maximality in languages with and without articles. *Journal of Semantics* 37(3). 311–366. <https://doi.org/10.1093/jos/ffaa002>.
- Šimík, Radek & Christoph Demian. 2021. Uniqueness and maximality in German and Polish: A production experiment. In Mojmir Dočekal & Marcin Wągiel (eds.), *Formal approaches to number in Slavic and beyond*, 149–171. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.5082462>.
- Sorace, Antonella. 2000. Introduction: Syntactic optionality in non-native grammars. *Second language research* 16(2). 93–102. <http://www.jstor.org/stable/43103572>.
- Stankov, Valentin. 1995. Semantični osobnosti na kategorijata neopredelenost na imenata v bálgarskija ezik. In Valentin Stankov (ed.), *Problemi na gramatičnata sistema na bálgarskija ezik*, 87–150. Sofia: Akademichno izdatelstvo “Prof. Marin Drinov”.
- Topolińska, Zuzanna. 1984. Składnia grupy imiennej. In Zuzanna Topolińska (ed.), *Gramatyka współczesnego języka polskiego. Składnia*, 301–386. Warszawa: Państwowe Wydawnictwo Naukowe.

APPENDIX

Example items in Bulgarian (a), Polish (b), and Russian (c):

- 1 EXISTENTIAL_MOD
 - a. V edno selo živeela _____ starica, kojato postojanno sábirala koreni i lista. Nikoj ne znael za kakvo sa ì služili.
 - b. W wiosce żyła _____ staruszka, która zbierała korzenie i liście. Nikt nie rozumiał, do czego jej służyły.
 - c. Žila-była v sele _____ staruxa, kotoraja sobirala koren’ja i list’ja. Nikto ne ponimal, dlja čego oni byli nužny.
‘In a village, there lived an old lady, who gathered roots and leaves. Nobody understood why they were needed.’
- 2 EXISTENTIAL
 - a. Imalo edno vreme v edin zamák _____ kralica. Tja bila gruba s vsički.
 - b. Dawno, dawno temu w zamku żyła _____ królowa. Była niesympatyczna dla wszystkich.
 - c. Davnym-davno v zamke žila _____ koroleva. Ona była nevežliva so vsemi.
‘A long time ago there lived a queen in a castle. She was impolite to all.’
- 3 POSTVERBAL_MOD
 - a. Prijatelite ti bjaxa izključitelno razválnuvani. Zašto? -Na koncerta pja _____ pevec, kojto naskoro stana izvesten.

- b. Tvoi przyjaciele byli bardzo poruszeni. Dlaczego? -Na koncercie występował _____ piosenkarz, który niedawno stał się sławny.
- c. Tvoi druž'ja čezvyčajno volnovalis'. Počemu? -Na koncerte vystupal _____ pevec, kotoryj nedavno stal izvestnym.
'Your friends were extremely worried. Why? -A singer who recently became famous performed in the concert.'

4 POSTVERBAL

- a. Kotkata ti beše pritesnena. Zašto? -Navn laeše _____ kuče.
- b. Twój kot był niespokojny. Dlaczego? -Na zewnątrz czekał _____ pies.
- c. Tvoja koška volnovalas'. Počemu? -Na ulice lajala _____ sobaka.
'Your cat was worried. Why? -In the street, a dog was barking.'

5 PREVERBAL_MOD

- a. Pred carkvata se bjaxa sábrali kup xora. Kakvo se slučvaše? - _____ sveštenik, kojto beše došál ot Vatikana, propovjadvaše.
- b. Przed kościołem zebrała się grupa ludzi. Co się tam działo? - _____ ksiądz, który przyjechał z Watykanu, wygłaszał kazanie.
- c. Pered cerkov'ju sobralas' kuča ludej. Čto tam bylo? - _____ svjaščennik, kotoryj prijexal iz Vatikana, propovedoval.
'A crowd gathered in front of the church. What was happening there? -A priest who came from the Vatican was preaching.'

6 PREVERBAL

- a. Vsički v ofisa gledaxa prez prozoreca. Zašto? - _____ čovek ležeše po sredata na ulicata.
- b. Wszyscy w biurze wyglądali przez okno. Dlaczego? - _____ mężczyzna leżał na środku drogi.
- c. Vse v ofise smotreli v okno. Počemu? - _____ mužik leżał posredi dorogi.
'Everybody in the office was looking outside the window. Why? -A man was lying in the middle of the road.'