The Grammaticalization of Numeral 'One' in Slavic: From Quantification to (Non-) Referentiality*

Daria Seres
University of Graz
daria.seres-guillen@uni-graz.at
https://orcid.org/0000-0002-9044-8516

Abstract

The article presents a corpus-based investigation of the distribution and interpretation of *'one'* + *N* combination in six Slavic languages (Russian, Ukrainian, Czech, Serbian, Macedonian, and Bulgarian). The study aims at determining whether there is an ongoing process of grammaticalization of the numeral 'one' in Slavic, and, if so, at what stage of grammaticalization is each language situated in the synchrony. In order to categorize the relevant uses of 'one' a system of semantic annotation is devised. The results of the study show different degree of grammaticalization of 'one' in the languages under study, with Russian and Ukrainian being in the beginning of the path (with occasional referential uses attested), Czech and Serbian being in between (with a significant frequency of referential uses), and Bulgarian and Macedonian being the most advanced on the scale (with a few non-referential uses). None of the languages has a fully grammaticalized indefinite article.

Keywords: grammaticalization; indefinite article; referentiality; Slavic languages

Resum. *La gramaticalització del numeral 'u' en les llengües eslaves: de la quantificació a la (no)referencialitat*

L'article presenta una investigació basada en el corpus de la distribució i interpretació de la combinació *'u'* + *N* en sis llengües eslaves (rus, ucraïnès, txec, serbi, macedoni i búlgar). L'estudi té com a objectiu determinar si hi ha un procés continu de gramaticalització del numeral 'u' en les llengües eslaves, i, en cas afirmatiu, en quina fase de gramaticalització es troba cada llengua en la sincronia. Per tal de categoritzar els usos rellevants del lexema 'u' es dissenya un sistema d'anotació semàntica. Els resultats de l'estudi mostren un grau diferent de gramaticalització d''u' en les llengües objecte d'estudi, amb el rus i l'ucraïnès al principi del procés (amb usos referencials ocasionals),

el txec i el serbi en un estadi intermedi (amb una freqüència significativa d'usos referencials), i el búlgar i el macedoni com els més avançats a l'escala (amb uns quants usos no referencials). Cap de les llengües té un article indefinit totalment gramaticalitzat.

Paraules clau: gramaticalització; article indefinit; referencialitat; llengües eslaves

Table of Contents

1. Introduction

The grammaticalization of numeral 'one' as an indefinite article is a widely known phenomenon found in many languages, including typologically different ones. For instance, this process has been attested in Indo-European languages, such as Romance languages or Germanic (apart from Icelandic, which lacks an indefinite article), in Dravidian languages, such as Tamil or Kannada (as reported in Heine 1997), in Turkic languages, such as Turkish or Uzbek (for the latter, see von Heusinger & Klein 2013), or in Basque. Cross-linguistically, throughout its diachronic development numeral 'one' may acquire a fascinating variety of new meaning and functions, such as a marker of specificity, referentiality, indefiniteness, genericity, or individuation, eventually evolving into an indefinite article (Heine 1997; Bybee 2002; Heine & Kuteva 2006; van Gelderen 2011, i.a.).

The current paper focuses on the grammaticalization of numeral 'one' as an indefiniteness marker / article in Slavic languages. This group of languages has not been thoroughly studied in this respect, as, according to typological descriptions (e.g., Dryer 2013b), they lack indefinite articles in their grammatical system. Nonetheless, numeral 'one' may have some of the functions, similar to those of an indefinite article, for example, marking of a newly introduced referent or marking a referent known to the speaker but unknown to the listener.

The present study aims at determining whether there is ongoing grammaticalization of numeral 'one' in Slavic languages, and if so, how advanced it is at the synchronic stage of the evolution of each language. To achieve this, an empirical corpus-based study of the distribution and interpretation of *'one' + N* construction was carried out on six Slavic languages. Numeral 'one' has the following base forms (masculine, singular, nominative case) in the languages under study: Russian *odin*, Ukrainian *odyn*, Czech *jeden*, Serbian *jedan*, Macedonian *eden* and Bulgarian *edin*, as illustrated in (1a-c).[1] The base form can be inflected for gender,

---

1.   Hereinafter in examples, ISO 639-3 codes are used to refer to the languages under study: RUS = Russian, UKR = Ukrainian, CES = Czech, HBS = Serbian, MKD = Macedonian, BUL = Bulgarian.

number, and case, except for Macedonian and Bulgarian, which do not display a complete paradigm of nominal case morphology in the synchrony. It should be noted, however, that these two languages have a fully developed definite article, unlike Russian, Ukrainian, Czech, and Serbian, which are considered to be completely articleless (Dryer 2013a, b).

(1) *(Beginning of a fairy-tale)*
    a. RUS Byla   na svete *odna   tëtja*.
               was.F.SG on world one F.SG. auntie
    b. UKR Bula   na sviti *odna   t'otja*.
               was.F.SG on world one F.SG. auntie
    c. CES Byla   jednou *jedna   teta*.
               was.F.SG once   one F.SG. auntie
    d. HBS Bila   jednom *jedna   teta*.
               was.F.SG once   one F.SG. auntie
    e. MKD Imaše   na zemjava   *edna   teta*.
               had.3SG on earth.DEF.PROX one F.SG. auntie
    f. BUL Imalo   edno vreme *edna   lelička*.
               had.3SG one time   one F.SG. auntie
               'Once there was *a certain auntie*.'

The main goals of this paper are as follows: (i) to define if there is an ongoing process of an indefinite article development in each Slavic language under study, (ii) to establish differences between typologically close languages with respect to the use of *'one'+N* construction, (iii) to confirm or disprove previous theoretical studies regarding the rise of an indefinite article in Slavic, using empirical data obtained through a corpus study method.

This article is organized in the following way: Section 2 is focused on the previous studies regarding the grammaticalization of numeral 'one' in different languages with a special focus on Slavic; Section 3 describes the empirical study, its methodology and the dataset; Section 4 explains the semantic annotation scheme, developed for this study; Section 5 presents the results of the study; in Section 6 these results are discussed; Section 7 concludes the paper and suggests further research steps.

2. Previous studies

2.1. The stages of grammaticalization of 'one'

Grammaticalization is generally understood as a process of diachronic development of a lexical item into a grammatical element. As a result of this process, the grammaticalized item normally undergoes a change in usage and distribution (Heine & Reh 1984), losing its lexical characteristics in favour of a functional/grammatical meaning. The item may also become phonologically reduced (e.g., the article 'a/an' vs numeral 'one' in English).

Throughout its grammaticalization, numeral 'one' gave rise to a new category – an indefinite article. As the evidence from many languages shows, the path from a lexical item that encodes cardinality (a singularity meaning) to an item that individuates a referent (an indefinite article) includes intermediate stages, when this item asserts the existence of a referent and encodes a property of speaker identifiably / specificity (an indefinite determiner).

According to Givón (1981), Heine (1997), among others, grammaticalization of 'one' cross-linguistically proceeds in the stages, as summarised in (2).

(2)   (i)   numeral ('one' has a quantitative function, expressing the cardinality of "1")

(ii)   presentative marker ('one' introduces a new referent which is taken up in the subsequent discourse; the referent is presumably unknown to the listener and known to the speaker)

(iii) specific marker ('one' introduces a new referent which does not have to be taken up in the subsequent discourse; the referent is known to the speaker but not to the listener)

(iv) non-specific marker ('one' introduces a referent which does not have to be known either to the speaker or to the hearer)

(v)   generalised article ('one' does not necessarily introduce a referent, it may be used generically, in predicative constructions, and in the scope of negation and modal operators).

<div align="right">adapted from Geist (2013: 5), Hwascz & Kedzierska (2018a: 96),
Molinari (2023: 4)</div>

Stage (i) represents a non-grammaticalized lexical item (a numeral), stages (ii)-(iv) correspond to an indefinite determiner, and stage (v) is a fully grammaticalized indefinite article.[2] According to Givón (1981), throughout their grammaticalization into indefinite articles, numerals undergo a process of *semantic bleaching*, which can be represented as the following scheme, containing three stages: quantification > referentiality > non-referentiality.

Notice that each stage of grammaticalization in (2) is assumed to incorporate the previous one, and thus, this process is considered unidirectional. That is, if a language has reached a certain stage, the uses of 'one' at all the previous stages will be attested in that language. The current paper aims at determining the stage of grammaticalization of 'one' in each of the Slavic languages under study, assessing its functions and the frequency of distribution.

2.2. The rise of an indefinite article in Slavic

There has been extensive research on the development of 'one' into an indefinite article in Romance languages (Laca & Tasmowski 1996; Ihsane 2008;

---

2.   Notice that some authors, for instance, Geist (2013), consider stage (iv) to correspond to an indefinite article, conflating non-specificity and non-referentiality.

Leonetti 2012; Pozas Loyo 2022, among many others) and in Germanic languages (Lockwood 1968; Heine 1997; Heine & Kuteva 2006; Leiss 2007; Crisma 2015, among many others). In some of these languages the form of the numeral 'one' is identical with the form of the indefinite article (e.g., Catalan, French, Italian, German, Swedish, etc.), in others, a new form has developed for the article, in most cases by means of phonological reduction (e.g., English, Dutch, Danish, etc.). Unlike the grammaticalization of 'one' in Germanic and Romance, a similar process in Slavic languages has not been thoroughly explored.

Apparently, in all the Slavic languages under study 'one' functions as a numeral, but its further functioning has not been described in detail in the linguistic literature or the assumptions regarding the use of this element have been based on limited empirical data. In what follows, I present a brief overview of theoretical accounts of the use of 'one' in Slavic.

In Russian, non-numeral uses of 'one' have been reported in the literature. Russian Academic Grammar (Švedova et al. 1980) calls the adnominal 'one' a *counting-pronominal,* while in the Russian Corpus Grammar (Padučeva 2016, 2017), it is defined as a *weak determinacy pronoun. 'One' + N* construction, according to Padučeva (2016), may be found in presentative contexts, as illustrated in (3).

(3) V molodosti poljubil   *odnu   ženščinu*,   xotel   na nej ženit'sja,
     in youth   fell.in.love one.ACC.F woman.ACC wanted on her marry,
     no ona otvetila   otkazom.
     but she answered refusal.INSTR
     'In my youth, I met *a woman*, wanted to marry her, but she declined.'
        from Russian National Corpus [Ю. Безелянский. В садах любви (1993)]

A non-presentative use of *odin* is also possible, as shown in (4). In this case, the *'one' + N* combination denotes a referent which is assumed not to be known to the listener.

(4) V èto vremja načalas' russko-japonskaja vojna. *Odin   čelovek*
     in this time   began   Russian-Japanese war.   one.NOM man.NOM
     predložil mne zanjatsja   postavkami.
     offered   me   occupy.myself supplies.INSTR
     'At that time the Russian-Japanese war began. *One man* suggested that I should occupy myself with supplies.'
     from the Russian National Corpus [А. Н. Толстой. Черная пятница (1924)]

'One' + N combinations are also found in Russian in situations that convey *epistemic inequality* (term by Padučeva 2016), when the speaker knows the referent but intentionally does not name it (which may even irritate the listener), as illustrated in (5).

(5) A: Èto… prosili   peredat…
         this   asked.PL.IPF hand

B: Kto?
who

A: *Odin čelovek*.
one.NOM man.NOM

B: Ja tebe pokažu *odnogo čeloveka*! Ja tebe ne odnogo,
I you.DAT show.1SG.PF one.ACC man.ACC I you.DAT not one.ACC
a mnogo čelovek pokažu! V tjurˈmax i zonax
but many people.GEN show.1SG.PF in prisons and labour.camps
ix, kak peska v pustyne!
them.GEN as sand.GEN in desert
'A: So… They asked me to hand it to you. B: Who? A: *One man*. B:
I gonna show you *one man*! I gonna show you not one man, but many
people. In prisons and labour camps there are as many of them as grains
of sand in the desert.'

from Russian National Corpus
[М. Гиголашвили. Чертово колесо (2007)][3]

In the absence of referential intent, as illustrated in (5), we can say that *odin* expresses anti-specificity. Thus, *odin* can be used both in association with a specific and an anti-specific reading, both of which presuppose the speaker's knowledge of the referent.[4]

It is also important to point out that 'one' is considered to be a non-obligatory specificity marker in Russian (Padučeva 1985, 2017; Ionin 2013; Borik 2016, i.a.).

In Ukrainian, only numeral use of 'one' has been reported to be possible, while the use in the presentative function (6a) or as a specific (6b) or non-specific (6c) indefinite is not allowed. A bare nominal is used in all these cases.

(6) a. Kolys' žyv sobi *učytel'*.
once.upon.a.time lived REFL teacher.NOM
'Once upon a time there lived *a teacher*.'

b. Odnoho razu *učytelja* bulo bačeno v policejskij diljanci.
one time teacher.ACC was seen in police office
'One day there was *a teacher* at the police office.'

c. Ja xoču maty *druga*.
I want have friend.ACC
'I want to have *a friend*.'

---

3. Notice that in this dialogue 'one man' in its anti-specific use gets reinterpreted as a numeral, that is, 'not one man but many people'.
4. For anti-specificity expressed by indefinites see Jayez & Tovena (2013), Etxeberria & Giannakidou (2017), and Espinal & Cyrino (2021), among others.

These are only a few examples that can be found in the literature, both descriptive and formal. There is no mention of 'one' among indefinite determiners in Pugh & Press (1999)'s "Comprehensive Grammar of Ukrainian". In general, among the six languages under study, Ukrainian can be considered the least studied. The current paper aims at filling in this gap at least partly.

In Bulgarian, 'one' is widely used as an indefinite determiner for both specific and non-specific referents (Geist 2013; Gorishneva 2013, 2016). It has even been claimed to function as an article-like element, able to express generic reference (Topolinjska 2009), as illustrated (7) (example from Geist 2013: 142).

(7) *(*Edin*) *džentâlmen* vinagi otvarja vrata na   damite.
       one.M gentleman always opens   doors PREP ladies.DEF
       '*A gentleman* always opens doors for ladies.'

With reference to 'one', Dimitrova-Vulchanova & Giusti (1998: 357) would rather speak of *an indefinite adjective*, since it agrees in gender and number with the noun it modifies and can co-occur with and actually carry the definite article, as illustrated in Tomić (2006: 55, fn. 18), see (8). When *edin* co-occurs with the definite article, it has a meaning of 'one of the given set'.

(8) *Edna-ta žena*   dojde.
       one.F.DEF woman came
       '*One of the women* came.'

In Macedonian, like in Bulgarian, 'one' can express a specific and a non-specific reference, but supposedly, it is less grammaticalized than in Bulgarian (Belaj & Matovac 2015), as it cannot be used generically (Tomić 2006: 58, fn. 23). Topolinjska (2009: 181), however, draws some examples that show the possible use of *eden* in the scope of a modal operator, which is characteristic of a generalised article, as illustrated in (9). So, determining whether it is more grammaticalized or less than in Bulgarian is not that straightforward. Weiss (2004), Tomić (2006), Topolinjska (2009), among others, consider *eden* an article-like element.

(9) a. Ova treba da   go   napravi *eden student*.
            this needs PART him do.3SG.PF one   student
            'This should be done by *a student*.'

     b. Bi možel da   doneseš   *edna igračka*.
            might   PART bring.2SG.PF one   toy
            'You could bring her *a toy*.'

In Czech, according to Cummings (1998), 'one' is a lexical quantifier that expresses specificity (speaker's knowledge of the referent), it is not obligatory but rather frequent especially for preverbal subjects, it "is approaching the status of an indefinite article" (Cummings 1998: 200). According to Heine & Kuteva

(2006: 123-124, ex. 3.42), *jeden* may be used adnominally when topic continuity is observed, as shown in (10).

(10) Včera   jsem potkal *jednoho Afričana*,   a   on mi řekl…
Yesterday am   met   one.ACC African.ACC and he me told
'Yesterday I met *an African* and he told me…'

In Serbo-Croatian, as stated in Kunzmann-Müller (1994: 91), there is no such morphological category as an article, but numeral 'one' can be used as a lexical means to express indefiniteness.[5] However, *jedan* is never obligatory, in contrast to languages with full-fledged articles, such as English or French. Belaj & Matovac (2015: 9, ex. 4d) report some uses of 'one' in predicative constructions, which, most probably, have an expressive / intensifying function, as illustrated in (11). Similar uses have been attested in Bulgarian, see Gorishneva (2016), Hwascz & Kedzierska (2018a), but they are rather infrequent.

(11) Ti   si   *jedna velika budala*.
you are one.F big   fool
'You are *a big fool*.'

Hwaszcz & Kedzierska (2018a) present a cross-linguistic study of the rise of an indefinite article in Slavic languages. Their assumptions are based on theoretical literature and have not been empirically checked. According to them, the grammaticalization of 'one', as presented in (2), in the languages covered by the current study has reached the following stages: (i) numeral in Ukrainian, (ii) presentative marker in Russian and Czech, (iii) specificity marker in Serbian, (iii) - (iv) specificity and non-specificity marker in Bulgarian and Macedonian. Thus, in none of the languages 'one' has reached the final stage of grammaticalization (a generalized article). As stated in the introduction to this paper, one of the goals of the current corpus-based investigation is to collect empirical data so as to check if the theoretical assumptions on the grammaticalization of 'one' in Slavic hold or whether they need to be updated.

Previous corpus-based studies on Slavic languages are rather scarce. Gorishneva (2016) compares the distribution of 'one'-marked vs. bare NPs in several types of Bulgarian corpora and concludes that the former ones are more suitable for establishing novel discourse referents (as they have a token-denotation), while the latter are found in generic contexts and with non-discourse-salient referents (they have a type-denotation).[6]

---

5. The language is called *Serbo-Croatian* here the way it is named in the cited literature. Alternatively, the language (or the South Slavic dialect continuum) may be called *Bosnian-Croatian-Serbian* (BCS) or *Bosnian-Croatian-Montenegrin-Serbian* (BCMS). In the current paper, I am using the term *Serbian*, as this is the way it is named in ParaSol corpus, from which I take my data.
6. For type-token distinction see Peirce (1906).

The only Slavic language that Hwaszcz & Kedzierska (2018b) investigate based on corpus data is Polish, where 'one' is found to be at stage (iii) of grammaticalization, according to the stages presented in (2). That means that 'one' may be used as a marker of specificity. This result is (partly) confirmed by the parallel corpus study of Borik et al. (*to appear*) of referential expressions, including *'one' + N*, in Russian, Polish and Macedonian. This study shows that, indeed, 'one' may mark specific indefinites in Polish, but the overall distribution of *'one' + N* in Polish patterns together with Russian, which goes against the assumption presented in Hwaszcz & Kedzierska (2018a) that in Russian 'one' may only be used in presentative contexts, that is, it is less grammaticalized than in Polish. On the whole, the study by Borik et al. (*to appear*) revealed that both Russian and Polish have a small overall number of occurrences of *'one' + N* (as compared to bare nominals), while Macedonian heavily relies on this structure in indefiniteness contexts. It should be noted, however, that the study of Borik et al. (*to appear*), as acknowledged by the authors themselves, was performed on a relatively small dataset (*n*=82 for translations of the English *a + N* construction), but it already highlights the important differences between some of the Slavic languages. The current study was initially inspired by the findings of Borik et al. (*to appear*) and motivated by the need to create a larger dataset with more data points and more languages that will enable a more nuanced analysis of *'one' + N* in Slavic.

3. Parallel corpus study

3.1. Research questions

The parallel corpus study presented in the current paper aims at gathering empirical data from six Slavic languages in order to analyze the stage of grammaticalization of 'one' in each language. It aspires to answer the following research questions: 1) What meanings can adnominal 'one' express in the Slavic languages under study? 2) What is the status of *'one' + N* construction in each language?

3.2. Data and Methodology

The data for the study is taken from ParaSol corpus, whose name stands for *parallel aligned corpus of Slavic and other Languages* (von Waldenfels & Meyer 2006-).[7] This is a corpus that features almost all major modern Slavic literary languages, and its focus is on texts translated into many languages (not just pairs of languages), that is why it mainly consists of 20th century prose with translations published in the second half of the century. Another important strategy used in this corpus is to include texts translated from different languages, however, most translations from Slavic languages are translations from lager languages, such as Russian or Polish. ParaSol contains around 40 literary texts, which in this work are referred to as *subcorpora*, and their parallel translations into several languages, from 2 to 28. The

---

7.   <https://www.parasolcorpus.org/Ursynow/>.

overall number of tokens for the languages under study is as follows: 3637357 for Russian, 1017057 for Ukrainian, 1039148 for Macedonian, 2002697 for Bulgarian, 1629868 for Czech, and 1317484 for Serbian (von Waldenfels 2012: 293-295).

ParaSol includes linguistic annotation, such as lemmatisation and POS-tagging, which made querying for *'one'* + *N* construction possible. For the purposes of the current study, the search was performed on four subcorpora (i.e. four novels): three subcorpora of texts originally written in Slavic articleless languages (M. Bulgakov *Master and Margarita* in Russian, I. Andrić *The Bridge on the Drina* in Serbian and M. Kundera *The Unbearable Lightness of Being* in Czech) and one English subcorpus (J. K. Rowling *Harry Potter and the Philosopher's Stone*).[8] The use of several subcorpora was necessary to guarantee that the results do not depend on the characteristics of the source text (the language of the original and its grammatical properties, author's style, register, etc.) but rather represent a general trend in the grammar of a given language.[9] Translations from English were used as control to make sure that the occurrence of *'one'* + *N* in each language does not depend on the presence or absence of the indefinite article in the original text.

The query for *'one'* + *N* in every language under study was limited by 100 hits per subcorpus, which means that for each language the maximum of 400 data points was expected to be collected (except for Serbian, which had the expected maximum of 300 data points as ParaSol does not contain the translation of Kundera's novel into Serbian).[10] The data points extracted from the corpora were manually annotated for the meaning of *'one'* + *N*, according to the annotation scheme presented in Section 4.

As mentioned in Section 2.2, the starting point for the present paper was Borik et al. (*to appear*)'s study, which was performed using Translation Mining approach to parallel corpora (Bremmers et al. 2021; van der Klis et al. 2022; Liu et al. 2023a; Liu et al. 2023b, i.a). Translation Mining is a methodology based on exploring the grammatical forms which different languages use in one and the same context.

Even though the methodology of the current study is different: only one form (*'one'* + *N* construction) is investigated in different contexts in each language and this form is annotated for its meaning, some of the assumptions of Translation Mining approach are still important for this study. Namely, the assumptions that the meaning of a source text is kept constant in its translations, and that a certain grammatical form is chosen by the translator (a competent native speaker) to render

---

8. ParaSol does not contain original texts in Ukrainian, Bulgarian, and Macedonian with parallel translations into several other Slavic languages, that is why only three languages (Russian, Serbian, and Czech) were used as original Slavic sources.

9. Subcorpus-specific results for each language do not differ significantly (see Appendix), except for Bulgarian and Macedonian when the original text was in English: the number of referential uses of *'one'* + *N* was significantly higher, which may indicate that these uses are similar to the uses of the indefinite article in English, but the comparative analysis of parallel translations is out of the scope of this paper.

10. As the collected data show, not every language reached the maximum of 100 hits per subcorpus for *'one'* + *N* construction.

the meaning of the original as closely as possible and reflects grammatical similarities and difference of the respective languages. Notice that most of the data in the current study come from translated texts.

4. Annotation

An important part of the current investigation was to devise a system of semantic annotation in order to define the uses of *'one' + N* in each language, thus, determining the stage of grammaticalization of 'one' into an indefinite article. The human annotator (who is also the author of this article) had to rely on formal criteria to define the meaning of the given construction, but also sometimes on the context, checking the original text, from which the fragment was translated. This annotation system may not be ideal, and the labelling is not very fine-grained (one label may cover several subcategories of uses), but it serves its purpose, distinguishing between some of the most common uses of 'one' in the languages under study and representing the grammaticalization path from quantification to (non-)referentiality.

Notice that the annotation scheme for the uses of 'one' does not coincide with the stages of grammaticalization of 'one', as introduced in (2), and, thus, should not be regarded as a linear representation of subsequent stages, rather as a mere collection of labels. The empirical study showed that some of the uses of *'one' + N* deviate from the grammaticalization path, while some other uses cover more than one stage of grammaticalization.

As a result of this exploratory study, the following uses of *'one' + N* construction in each language were singled out:

1. NUMERAL
2. DERIVED meaning or IDIOMATIC uses
3. ONE + OTHER
4. PRESENTATIVE or SPECIFIC marker
5. NON-SPECIFIC, GENERIC marker or PREDICATIVE use.

The criteria for assigning each label are explained below in Subsections 4.1-4.5.

4.1. 'One' as a NUMERAL

The first task at annotating *'one' + N* is to distinguish between the numeral and non-numeral use. There are several tests to do it.

'One' is labelled as a numeral, when the cardinality meaning is at issue. That is, 'one' has a quantification interpretation, expressing singularity. It ranges over a set containing a single element, which leads to a scalar implicature (Dayal 2004; Krifka 2004, i.a.), as demonstrated in (12).[11]

---

11. All the examples hereinafter are taken from ParaSol corpus. The language is indicated in front of each example. The examples are not glossed as they are quite lengthy, the relevant expressions in the original language and their correspondences in the translation are given in italics.

(12) RUS "Potrebovat', potrebovat', moja donna, - otvečal Voland ponimajušče
          ulybajas', - potrebovat' *odnoj vešči*!"
          '"Demand, demand, my donna," Woland replied, smiling knowingly,
          "You may demand *one thing*."
          Interpretation: 'You may not demand more than one thing.'

Another hallmark of a numeral is an overt contrast in cardinality (Geist 2013),
which stems from a pragmatic relation of 'one' to alternative number words (Krifka
2004), as illustrated in (13).

(13) RUS V čisle pročego bylo […] *dva obeščanija* pokončit' žizn' samoubijstvom
          i *odno priznanije* v tajnoj beremennosti.
          'Among others there were […] *two vows* to end life by suicide and *one
          confession* of secret pregnancy.'

'One' as a numeral, but not as an indefinite marker, is compatible with restric-
tive focus particles and numeral modifiers (Geist 2013), such as 'exactly one', 'at
least one', 'one more', 'only one', etc.

(14) UKR Ron ne mih zrozumyty, ščo cikavoho v hri, de *til'ki odyn m'jač* i ne
          možna litaty.
          'Ron couldn't see what was exciting about a game with *only one ball*
          where no one was allowed to fly.'

Unlike an indefinite marker, 'one' as a numeral may be used distributively, as
shown in (15).

(15) MKD Sekoja godina gospodinot pripeduva *eden bal*.
          'Messire gives *one ball* annually.'
          Lit. 'Every year Messire gives *one ball*.'

Numerals can also be combined with measure units (which may not always be
conventional): 'one pound', 'one kilometre', 'one head higher', etc., time expres-
sions: 'for one hour', 'in one week', etc. and other expression, typically used with
numerals, for instance, 'one point', as illustrated in (16).[12]

(16) HBS A zbog tebe, Poteru, Grifindor gubi *jedan poen*!
          'And *a point* will be taken from Gryffindor House for your cheek Potter.'

---

12. In the corpora of Czech, Russian, Serbian, and Ukrainian (but not Bulgarian or Macedonian), NPs
    which expressed the cardinality of "1" sometimes appear as bare (this is a phenomenon worth
    exploring but it is out of the scope of the current paper). Notice that in English it is possible to use
    an indefinite article with a cardinal meaning.

    (i) a. MKD Dajte mi *edna nedela* da razmislam. ['one' + N]
        b. CES Dejte mi *týden* na rozmyšlenou. [N]
                'Give me *a week* to think it over.'

It should be acknowledged that sometimes the difference between a numeral and an indefinite marker is impossible to establish, as the context does not give enough information and there is none of the above-mentioned formal characteristics of numerals. The interpretation of *'one'* + *N* construction in sentences, such as (17), is ambiguous. It could have been disambiguated in speech by means of prosody: numerals are generally stressed, while indefinite markers are not, as illustrated in (17b) and (17c). However, obviously this is not possible with corpus data.

(17) RUS a. Šum imeet *odno preimuščestvo*. V něm propadajut slova.
         'Noise has *one advantage*. It drowns out words.'

     b. Šum imeet *ODNO preimuščestvo*. V něm propadajut slova.
         'Noise has *(only) one advantage*. It drowns out words.'

     c. Šum imeet *odno PREIMUŠČESTVO*. V něm propadajut slova.
         'Noise has *a certain advantage*. It drowns out words.'

In such ambiguous cases, 'one' is annotated as SPECIFIC and considered to be referential, that is, as in (17c). And still this label is used less often than the label NUMERAL.

4.2. DERIVED meaning or IDIOMATIC uses of 'one'

This label is used for constructions that do not express cardinality directly, but they are still related to the core meaning of 'one'. Gorishneva (2016: 146) calls them "semantic derivations of singularity" and distinguishes the following three types:

1) Event-related singularity, which may specify the number of occurrences of an event: UKR *odnoho dnja,* HBS *jednog dana*, CES *jednoho dne* 'one day', RUS *v odin prekransnyj den'*, 'once upon a time', BUL *edna nošč* 'one night', *edna nedelja* 'one Sunday'; or the manner of an event: RUS *odim duxom* 'at one gulp, lit. in one spirit', MKD *so eden zbor*, BUL *s edna duma* 'in short, lit. in one word'. Some of these occurrences are traditionally classified as idiomatic expressions, that is, "frozen bits of complex syntax" (Nattinger 1980: 337).

2) Object-related singularity, which represents quantification over objects, for instance, in (18) the expression *v odnoj soročke* 'lit. in one shirt' means 'in nothing but a shirt', 'only in a shirt', that is, its meaning is similar to a restrictive focus particle.[13]

---

13. An anonymous reviewer correctly points out that the *'one'* + *N* expression in (18) can be pluralized, providing the following example:

    (i) RUS Dvoe detej – mal'čik i devočka – v *odnix rubaškax* i bosye, šli pozadi cerkovnoj tolpy […]
         'Two kids – a boy and a girl – *in nothing but shirts* (lit. one.pl shirts) and barefoot, walked behind the church crowd […]'

                                                          (from A. Platonov *Rodina električestva*)

(18) RUS […] on uvidel na trotuare pod soboj damu v *odnoj soročke* […]
  '[…] he saw, just below him on the sidewalk, a lady in nothing but *a shirt* […]'

3) Distribution-related singularity, which expresses equality of tokens 'one and the same, selfsame' or equality of types 'of the same type', as illustrated in (19).[14]

(19) MKD […] nevozmožnost da se živee *vo eden stan* so banditi.
  '[…] the impossibility of living in *the same apartment* with bandits.'

It should be noted that derived and idiomatic uses of 'one' do not constitute a stage of grammaticalization; they can be viewed as deviations from the path.

4.3. ONE + OTHER

This label is used in presence of two disjoint referents: 'one + the other' or 'one + one more' / 'one + a different one'. This difference in the interpretation is not relevant for the annotation.[15] This use of 'one' is singled out because the referent of *'one' + N* belongs to a set, whose cardinality is either two or more than two. It was experimentally shown for Russian by Seres et al. (2023) that the presence of disjoint referents increases the speakers' preference for *'one' + N* over bare NPs. ONE + OTHER is often used for pair nouns, as shown in (18). Notice that Bulgarian (20b) and Macedonian use the definite marking on 'one' in such cases.

(20) a. CES *Jeden pól* protikladu byl pro něho pozitivní (světlo, teplo, jemnost, bytí), *druhý* negativní.

  b. BUL *Edini-ja poljus* v dvojkata prijemal za položitelen (svetlinata, toplinata, efirnostta, bitieto), a *drugi-ja* za otricatelen.
    'One half of the opposition he called positive (light, fineness, warmth, being), the other negative.'

Nonetheless, the meaning of *'one.pl' + N.pl* in this example remains the same as in (18), that is, 'nothing but a shirt', but the structure gets pluralized as it is used distributively w.r.t. the two referents – *boy* and *girl*. The use of 'one' is still not referential in this case; it rather functions as a restrictive modifier.

14. As an anonymous reviewer correctly predicts, (19) is translated into Serbian with an overt adjective *isti* 'same'.

  (i) […] nemogućnost da se živi u *istom stanu* sa banditima.

  However, this adjective expresses the equality of tokens and not referentiality. The Serbian translation confirms the correct interpretation (as distribution-related singularity) of 'one' in (19).

15. Ukrainian is the only language in this study that makes a lexical difference between 'one vs. the other' (ia) and 'one + a different one' (ib): *druhij* vs. *inšij*.

  (i) a. …ja z *odnoho boku*, vona z *druhoho*.
       '… I on one side, she on the other.'

    b. vin ide z *odnoho svjata* na *inše*.
       'like someone going from party to party.'

The interpretation of 'one' in this type of construction is still close to the numeral. It leads to a scalar implicature: 'one, and not more than one', and this is what matters for the current study.

4.4. 'One' as a SPECIFIC/PRESENTATIVE marker

This type of label is used when there is no cardinality inference and no scalar implicature can be triggered. Following Heine (1997), the distinction between a presentative and a specific marker as distinct stages of grammaticalization of 'one' has been present in the literature, as presented in (2). However, this distinction may not always be clear or relevant, or it may depend on the understanding of specificity (scopal, epistemic, referential, etc.).

In both cases a new referent presumably unknown to the listener but known to the speaker is introduced into discourse, which means that the referent is epistemically specific. In the case of a presentative marker, the referent is supposed to be taken up in the subsequent discourse, that is, it is characterised by discourse persistence / discourse prominence / salience; and in the case of a specific marker, the referent may but does not have to be discourse persistent. Thus, there is no difference in specificity when it is understood epistemically: the reference of an expression depends on the speaker's knowledge, that is, when the speaker has a particular referent in mind (von Heusinger 2002; Ihsane 2021), there may be difference in discourse prominence, but it is not an obligatory condition. That is why the two stages of grammaticalisation of 'one' are covered by one label: SPECIFIC/PRESENTATIVE.

Furthermore, this label includes cases when the speaker is aware of the existence of the referent, but not necessarily of its identity, as in (21). Arguably, the referent is still potentially identifiable.

(21) RUS *Odin filosof* odnaždy napisal obo mne, čto vsë, čto ja govorju, bezdoka-
      zatel'naja spekul'acija […]
      '*A philosopher* once wrote that everything in my work is unverifiable
      speculation[…]'

Summing up, what is important for assigning the label SPECIFIC/ PRESENTATIVE is the existence of the referent, but not the speaker's knowledge, which may not be available for the annotator, or the discourse prominence of the referent, which is not a necessary condition for existence.

4.5. Non-referential uses of 'one': NON-SPECIFIC, GENERIC, PREDICATIVE

The label NON-SPECIFIC, GENERIC, PREDICATIVE covers non-referential uses of 'one'. The occurrences of *'one' + N* with such uses were not many (see Section 5 for the results of the study), that is why they are under one label.

Non-specificity is understood here as non-refentiality, that is, the absence of existential commitment. Notice that this label does not depend on the speaker's knowledge (i.e., epistemic specificity), as something that an annotator may poten-

tially misinterpret. This label is given under two conditions: (i) *'one' + N* appears in a non-veridical context (in questions, conditionals, imperatives, in the scope of negation, modal verbs or non-veridical attitude verbs, such as, *want*, *insist*, *suggest*, *allow*, etc.) and (ii) the referent is not taken up in subsequent discourse.[16] Thus, the existence of the referent is not guaranteed, as in examples (22) and (23).

(22) MKD …sakam da imam *eden zmej*!
   'I'd like *a dragon*!'

(23) BUL Ama da ste viždali *edna žaba*?
   'Have you seen *a toad* at all?'

When *'one' + N* is used generically, it refers to a (typical) representative member of the kind denoted by the noun, but not to the kind itself.[17] There is no existential commitment of the referent. The generic interpretation is context-dependent and is considered accessible to the annotator. Such contexts usually express certain rules (including social norms or rules of everyday life), ability, (logical) necessity or disposition. Example (24) expresses a rule of everyday life.

(24) a. BUL Na izkâpanija Ivan Nikolaevič vednaga mu beše dadeno absoljutno vsičko, neobhodimo za *edin mâž* sled banja.

   b. MKD Na Ivan, koga se izbanja, mu beše dadeno se što mu treba na *eden maž* po kapanjeto.
   'A washed-up Ivan Nikolayevich was immediately given everything *a man* needs after his bath.'

Another non-referential use of *'one' + N* is identified by its syntactic function. This construction may be found in the predicate position, as illustrated in (25) for Macedonian.

(25) MKD Ti si, na primer, *eden izmannik*.
   'You, for instance, are *a liar*.'

The following section presents the distribution of the labels that has just been described in detail in the six Slavic languages under study.

5. Results

The empirical corpus-based study aimed at revealing the distribution of *'one' + N* combination in different uses in order to see the degree of grammaticalization of this 'one' in each of the six Slavic languages. Figure 1 presents this distribution.[18]

---

16. For more details on (non)veridicality, see Giannakidou (2013).
17. For more details on the distinction between generic and kind reference in languages with and without articles, see Borik & Espnial (2015, 2020), Seres (2020), among others.
18. See Appendix for the number of occurrences of *'one' + N* in all the languages under study, combined and with respect to each subcorpus.
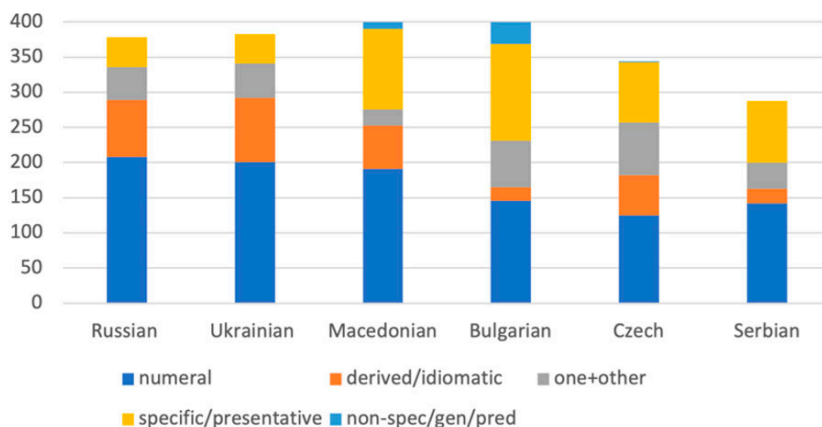
Figure 1. Uses of ONE+N across Slavic

First and foremost, it is important to point out that only Macedonian and Bulgarian reached the count of 100 hits per subcorpus, resulting in $n = 400$; in the other four languages, *'one' + N* was a bit less frequent (Russian, $n = 378$; Ukrainian, $n = 383$; Czech, $n = 322$; Serbian, $n = 288$).

In Russian and Ukrainian, in more than half of the cases (55% and 52.4%, respectively) 'one' is used as a numeral, while the referential uses (specific / presentative) are the lowest of all languages (11% in both languages).[19] The referential uses of *'one' + N* construction account for 34.5% of the uses in Bulgarian, 30% in Serbian, 28% in Macedonian, and 25% in Czech.

Fisher's Exact Test (FET), performed on pairs of languages in all possible combinations, confirmed that the difference in the distribution of referential *'one' + N* between Russian / Ukrainian and the other four languages is statistically significant ($p < 0.00001$), as well as the difference between Bulgarian and Czech ($p = 0.0051$).

Non-referential uses (non-specific / generic / predicative) were found in Bulgarian and Macedonian in 8% and 3% of cases, respectively, but not in other languages (except for 1 hit in Czech, which may be considered an outlier).

To sum up, the results show cross-linguistic variation in the distribution of *'one' + N* across the Slavic language subgroup, even though the languages under study are typologically and genetically close. In all the six languages, the most frequent use of 'one' is the expression of quantity, that is, 'one' functions as a numeral. This is an expected result as the expression of cardinality is the core role of 'one'. This function is the most prevalent in Russian and Ukrainian, while referential uses (presentative and specific) are the least frequent in these two

---

19. Notice that the uses annotated as ONE+OTHER and DERIVED/IDIOMATIC are not counted with reference to the grammaticalization scale of 'one', even though they are closer to quantificational use, as shown in Subsections 4.2 and 4.3.

languages and non-referential uses are absent. The frequency of occurrence of referential *'one' + N* is similar in Serbian, Czech, Bulgarian, and Macedonian, but non-referential uses are only found in the latter two, and they are rather sporadic.

6. Discussion

The results obtained in the corpus study may be interpreted as indicating different degrees of grammaticalization of 'one' in Slavic languages. The process of grammaticalization has started in all the six languages, as referential uses of 'one', both as a presentative and a specific marker, are attested in all of them.

The presence of referential uses of 'one' in Ukrainian goes against the previous theoretical assumptions about this language, which claim that only quantificational (numeral) uses are possible (e.g., Heine & Kuteva 2006; Hwaszcz & Kedzierska 2018a). Example in (26) shows a typical presentative context (an existential sentence) with discourse anaphora licencing, which introduces a novel referent that becomes the topic of the following discourse.

(26) UKR   U mully Ibrahima bula *odna vada*: vin syl'no zajikavsja.
               'Mullah Ibrahim had *one handicap*: he stuttered badly.'

Specific but not discourse-salient uses of 'one' are found in Ukrainian, Russian and Czech also go against the predictions in Hwaszcz & Kedzierska (2018a), as illustrated in (27), where the newly introduced referent has no discourse persistence.

(27) a. RUS Včera v restorane ja *odnomu tipu* po morde zasvetil', – mužestvenno
                priznalsja preobražënnyj poèt.

      b. UKR Včora v restorani ja *odnomu sub'jektovi* zasvityv u pyku, -mužn'o
                ziznavsja poet, svidomist' jakoho zaznala peretvoren'.

      c. CES Včera v restauraci jsem dal *jednomu chlapovi* po hubě, - přiznal se
                hrdinně jakoby vyměněný básník.
                '"I punched *a guy* in the face in a restaurant yesterday", the transformed poet bravely confessed.'

Even though all the six languages present evidence of the referential use of 'one', the difference in the frequency of its use in this function may give some insights into the stage of its grammaticalization in each language. In Russian and Ukrainian, showing the lowest frequency of referential *'one' + N* may be classified as languages where the grammaticalization is the least advanced.

With respect to referential uses, Serbian, Czech, Macedonian, and Bulgarian seem to occupy the same position on the grammaticalization scale. However, Macedonian and Bulgarian seem to have entered the next stage, where non-referential uses are also possible.

A more advanced stage of grammaticalization of 'one' in Bulgarian and Macedonian may be due to the presence of the definite article in the grammatical

system of these two languages. As has been stated in the literature (Moravcsik 1969; Heine 1997; Crisma 2011; Goldstein 2022, i.a.), the definite article often emerges in the system before the indefinite one, and the grammaticalization of an indefinite article is more likely to happen in languages that possess a definite article. Note, however, that the two articles develop independently, and there are quite a few examples of languages which only have an indefinite article (Dryer 2013b; Goldstein 2022).

It is important to highlight that non-referential uses of 'one' in Bulgarian and Macedonian are rather irregular and infrequent, which may indicate just the beginning of this stage of grammaticalization. Thus, the adnominal 'one' in these languages is still far from being considered a fully-fledged indefinite article.

Another observation, which goes in line with Belaj & Matovac (2015), concerns the difference in the generic use of *'one' + N* in Bulgarian and Macedonian. 'One' as a marker of generic reference is expected to be found in Bulgarian but not in Macedonian Indeed, in (28) and (29), it can be observed that while Bulgarian uses the adnominal 'one' (28a and 29a), Macedonian resorts to a bare nominal in (28b) (notice also that a bare NP is used in the English translation) and a definite NP in (29b).

(28) a. BUL […] za nas veličieto na *edin čovek* e v tova, če toj nosi svojata sâdba, kakto Atlant e nosel na ramenete si nebesnija svod.

    b. MKD […] goleminata na *čoveka* za nas leži vo toa što ja nosi svojata sudbina kako što Atlant go nosel na ramenja nebesniot svod.
'[…] we believe that the greatness of *man* stems from the fact that he bears his fate as Atlas bore the heavens on his shoulders.'

(29) a. BUL Kogato *edno obščestvo* e bogato, negovite členove ne sa prinudeni da se zanimaavat s fizičeski trud…

    b. MKD Koga e *opščestvo-to* bogato, luǵeto ne moraat da rabotat so race…
'When *a society* is rich, its people don't need to work with their hands…'

Nonetheless, generic uses of *'one' + N* are also attested in Macedonian, as was shown in example (24). Thus, the data do not clearly indicate whether 'one' is more grammaticalized in one language or another. It makes sense for now only to say that non-referential uses are attested in both languages, but this stage of grammaticalization has not been fully achieved yet.

The irregularities in the data and the differences in the frequency of occurrence of 'one' in different uses may also indicate the optionality of its use in a given function. Obviously, one of the limitations of corpus research is the absence of negative data. That is, it is impossible to assess the degree of optionality of a construction that is found in a corpus. However, it can be suggested that a lower frequency of use signals a higher degree of optionality.

The data obtained in the present study are compatible with the claims from theoretical literature on Russian that state that 'one' as an indefiniteness marker is

never obligatory in this language (Padučeva 1985, 2017; Ionin 2013; Borik 2016, i.a.). A similar frequency of occurrence of referential uses of 'one' in Ukrainian may suggest that this claim can be extended to this language as well. A similar hypothesis may be proposed for non-referential uses of 'one' in Bulgarian and Macedonian: a very low frequency of occurrence of 'one' in such contexts may indicate its total optionality.

Another limitation of this corpus study may concern the type of texts that we were used for the data collection. They are literary works which present a certain degree of language codification. Even though they contain both narration and dialogues, they cannot be considered representative of spoken language. Moreover, all the texts were written in the previous century (between 1930s and 2000s), so they cannot reflect the most recent changes in a language. Nonetheless, the data gathered from these corpora can reliably show general trends in the grammar of each language.

7. Conclusions

The empirical corpus study of the distribution and interpretation of *'one' + N* construction in six Slavic languages made it possible to collect robust data in order to determine the degree of grammaticalizing of 'one' in each of the languages. In all the languages under study 'one' can used not only as a numeral but also as an indefinite determiner, which asserts the existence of the referent. However, judging from the frequency of occurrence, in Russian and Ukrainian the process of grammaticalization of 'one' as an indefinite determiner is in its initial stage, while it is more advanced in Serbian and Czech. In Macedonian and Bulgarian, the incipient use of non-referential *'one' + N* is attested, which signals the beginning of the next stage of grammaticalization, but it is still far from being grammaticalized as a full-fledged indefinite article.

Corpus data can give a lot of insights about the stage of grammaticalization of 'one' in the languages under study, but the discovered tendencies need to be confirmed by other empirical methods (e.g., experimental studies with native speakers to build up critical mass). Further studies may also bring light to the question of optionality / obligatoriness of the use of grammatical structures, which cannot be reliably explored by means of a corpus study.

References

Belaj, Branimir & Matovac, Darko. 2015. On article-like use of indefinite determiners *jedan* and *neki* in Croatian and other Slavic languages. *Suvremena lingvistika* 79(1): 1-20.

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 & Espinal, Maria Teresa. 2015. Reference to kinds and to other generic expressions in Spanish: definiteness and number. *The Linguistic Review* 32(2): 167-225.
    <http://doi.org/10.1515/tlr-2014-0023>

Borik, Olga & Espinal, Maria Teresa. 2020. Numberless kinds: evidence from Russian. *Catalan Journal of linguistics* 19: 231-260.
<https://doi.org/10.5565/rev/catjl.299>

Borik, Olga, Le Bruyn, Bert, Liu, Jianan & Seres, Daria. (to appear). Bare nouns in Slavic and beyond. In Gehrke, Berit, Lenertová, Denisa Meyer, Roland, Seres, Daria, Szucsich, Luka & Zaleska, Joanna (eds.). *Advances in Formal Slavic Linguistics 2022*. Berlin: Language Science Press.

Bremmers, David, Liu, Jianan, van der Klis, Martijn & Le Bruyn, Bert. 2021. Translation mining: Definiteness across languages (A reply to Jenks 2018). *Linguistic Inquiry* 53(4): 735-752.
<https://doi.org/10.1162/ling_a_00423>

Bybee, Joan L. 2002. Cognitive processes in grammaticalisation. In Tomasello, Michael (ed.). *The New Psychology of Language*, vol. II, 145-167. New Jersey: Lawrence Erlbaum Associates Inc.

Crisma, Paola. 2011. The emergence of the definite article in English: A contact-induced change? In Sleeman, Petra & Perridon, Harry (eds.). *The noun phrase in Romance and Germanic: Structure, variation, and change*, 175-192. Amsterdam: John Benjamins.
<https://doi.org/10.1075/la.171.13cri>

Cummins, George. 1998. Indefiniteness in Czech. *Journal of Slavic Linguistics* 6(2): 171-203. <http://www.jstor.org/stable/24599695>

Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4): 393-450.
<https://doi.org/10.1023/B:LING.0000024420.80324.67>

Dimitrova-Vulchanova, Mila & Giusti, Giuliana. 1998. Fragments of Balkan Nominal Structure. In Alexiadou, Artemis & Wilder, Chris (eds.). *Possessors, Predicates and Movement in the Determiner Phrase*, 333-360. Amsterdam: Benjamins.
<https://doi.org/10.1075/la.22.12dim>

Dryer, Matthew S. 2013a. Definite articles. In Dryer, Matthew S. & Haspelmath, Martin (eds.). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/37>

Dryer, Matthew S. 2013b. Indefinite articles. In Dryer, Matthew S. & Haspelmath, Martin (eds.). *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/38>

Espinal, Maria Teresa & Cyrino, Sonia. 2021. A syntactically-driven approach to indefiniteness, specificity and anti-specificity in Romance. *Journal of Linguistics* 58(3): 535-570.
<https://doi.org/10.1017/S002222672100030X>

Etxeberria, Urtzi & Giannakidou, Anastasia. 2017. *A unified analysis of Spanish algún/ algunos: anti-specificity, plurality, and NP-anaphora*. Ms. <https://home.uchicago. edu/~giannaki/pubs/etxeberria.giannakidou.algunos.2017.sept.pdf>

Geist, Ljudmila. 2013. Bulgarian *edin*: The rise of an indefinite article. In *Formal Description of Slavic Languages: Proceedings of FDSL 9, Göttingen 2011*, 125-148. Frankfurt am Main: Peter Lang.

van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. New York: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199756056.001.0001>

Giannnakidou, Anastasia. 2013. (Non)veridicality, evaluation, and even actualization: evidence from subjunctive in relative clauses. In Taboada, Maite & Tvranc, Radoslava (eds.). Nonveridicality, Perspective, and Discourse Coherence. Leiden: Brill.
<https://doi.org/10.1163/9789004258174_001>

Givón, Talmy. 1981. On the development of the numeral 'one' as an indefinite marker. *Folia Linguistica Historica* 15(2-1): 35-54.
<https://doi.org/10.1515/flih.1981.2.1.35>

Gorishneva, Elena. 2013. Bare vs. non-bare nouns. Two kinds of indefinites in Bulgarian. In Kabatek, Johannes & Wall, Albert (eds.). *New Perspectives on Bare Noun Phrases in Romance and Beyond*, 301-328. John Benjamins.
<https://doi.org/10.1075/slcs.141.11gor>

Gorishneva, Elena. 2016. *The variety of functions of the numeral and indefinite marker 'one' in Bulgarian and Russian*. Herne: Gabriele Schäfer Verlag.

Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.
<https://doi.org/10.1093/oso/9780195102512.001.0001>

Heine, Bernd & Kuteva, Tania. 2006. *The Changing Languages of Europe*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199297337.001.0001>

Heine, Bernd & Reh, Mechthild. 1984. *Grammaticalisation and Reanalysis in African Languages*. Hamburg: Helmut Buske Verlag.

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>

von Heusinger, Klaus & Klein, Udo. 2013. The Distribution of two Indefinite Articles: The Case of Uzbek. In Ebert, Cornelia & Hinterwimmer, Stefan (eds.). *Different Kinds of Specificity Across Languages*. Studies in Linguistics and Philosophy, vol. 92. Dordrecht: Springer.
<https://doi.org/10.1007/978-94-007-5310-5_7>

Hwaszcz, Krzysztof & Kędzierska, Hanna. 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 & Kędzierska, Hanna. 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.18.007.9256>

Ihsane, Tabea. 2008. *The layered DP: From meaning of French indefinites*. Amsterdam: John Benjamins.
<https://doi.org/10.1075/la.124>

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>

Ionin, Tania. 2013. Pragmatic Variation among Specificity Markers. In Hinterwimmer, Stefan & Ebert, Cornelia (eds.). *Different Kinds of Specificity Across Languages*, 85-103. Dordrecht: Springer.
<https://doi.org/10.1007/978-94-007-5310-5_4>

Jayez, Jacques & Tovena, Lucia. 2013. Scenarios of equivalence: the case of quelque. In Hinterwimmer, Stefan & Ebert, Cornelia (eds.). *Different kinds of specificity across languages*, 177-207. Dordrecht: Springer.
<https://doi.org/10.1007/978-94-007-5310-5_8>

van der Klis, Martijn, Le Bruyn, Bert & de Swart, Henriëtte. 2022. A multilingual corpus study of the competition between past and perfect in narrative discourse. *Journal of Linguistics* 58(2): 423-457.
<https://doi.org/10.1017/S0022226721000244>

Krifka, Manfred. 2004. Bare NPs: Kind-referring, indefinites, both or neither? *Semantics and linguistic theory (SALT)* 13: 180-203.
<https://doi.org/10.3765/salt.v13i0.2880>

Kunzmann-Müller, Barbara. 1994. *Grammatikhandbuch des Kroatischen und Serbischen*. Frankfurt am Main/Berlin/Bern/New York/Paris/Wein: Peter Lang Verlag.
<https://doi.org/10.3726/978-3-653-01831-8>

Laca, Brenda & Tasmowski, Liliane 1996. Indéfini et quantification. *Recherches Linguistiques de Vincennes* 25: 107-128.

Leiss, Elisabeth. 2007. Covert patterns of definiteness/indefiniteness and aspectuality in Old Icelandic, Gothic and Old High German. In Stark, Elisabeth, Leiss, Elisabeth. & Abraham, Werner (eds.). *Nominal determination: Typology, context constraints and historical emergence*, 73-102. Amsterdam: John Benjamins.
<https://doi.org/10.1075/slcs.89.06lei>

Leonetti, Manuel. 2012. Indefiniteness and specificity. In Hualde, José Ignacio, Olarrea, Antxon & O'Rourke, Erin (eds.). *Handbook of Hispanic linguistics*, 285-306. Oxford: Wiley.
<https://doi.org/10.1002/9781118228098.ch15>

Liu, Jianan, Patil, Shravani, Seres, Daria, Borik, Olga & Le Bruyn, Bert. 2023a. 'Articleless' languages are not created equal. In Onoeva, Maria, Staňková, Anna & Šimík, Radek (eds.). *Proceedings of Sinn und Bedeutung* 27: 381-398.
<https://doi.org/10.18148/sub/2023.v27.1076>

Liu, Jianan, Patil, Shravani, Schurr, Hagay, Seres, Daria, Borik, Olga & Le Bruyn, Bert. 2023b. The theory of argument formation: between kinds and properties. In *Proceedings of SALT 33*, 298-319.
<https://doi.org/10.3765/bdzpks21>

Lockwood, William B. 1968. *Historical German syntax*. London: Oxford University Press.

Molinari, Luca. 2023. The Syntax of Bulgarian *edin* 'one'. *Journal of Slavic Linguistics* 31(1-2).

Moravcsik, Edith A. 1969. Determination. *Working Papers in Language Universals* 1: 63-98. Stanford: Stanford University.

Nattinger, James R. 1980. A lexical phrase grammar for ESL. *TESOL Quarterly* 14: 337-344.
<https://doi.org/10.2307/3586598>

Padučeva, Elena V. 1985. *Vyskazyvanie i ego sootnesënnost' s dejstvitel'nost'ju (referencial'nye aspekty semantiki mestoimenij)*. Moskva: Nauka.

Padučeva, Elena V. 2016. *Mestoimenija slaboj opredelënnosti (serija na koe-; serija na ne-; odin)*. Ms. Moscow. <http://rusgram.ru/>

Padučeva, Elena V. 2017. *Referencial'nyj status imennoj gruppy*. Ms. Moscow. <http://rusgram.ru/>

Peirce, Charles S. 1906. Prolegomena to an apology for pragmaticism. *Monist* 16: 492-546.
<https://doi.org/10.5840/monist190616436>

Pozas Loyo, Julia. 2022. Indefinite Articles in the Romance Languages. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
<https://doi.org/10.1093/acrefore/9780199384655.013.637>

Pugh, Stefan & Press, Ian. 1999. *Ukrainian: a comprehensive grammar*. London/New York: Routledge.

Seres, Daria. 2020. *The expression of genericity in languages with and without articles*. Barcelona: Universitat Autònoma de Barcelona (PhD Dissertation). <http://hdl.handle.net/10803/670853>

Seres, Daria, Borràs-Comes, Joan & Espinal, Maria Teresa. 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>

Švedova, Natalija et al. (eds.). 1980/1982. *Russkaja grammatika*. Moscow: Nauka.

Tomić, Olga. 2006. *Balkan Sprachbund morpho-syntactic features*. Dordrecht: Springer.
<https://doi.org/10.1007/1-4020-4488-7>

Topolinjska, Zuzanna. 2009. Definiteness (synchrony). In *Halbband 1*, 176-188. Berlin, New York: De Gruyter Mouton.
<https://doi.org/10.1515/9783110214475.1.3.176>

von Waldenfels, Ruprecht. 2012. ParaSol: Introduction to a Slavic Parallel Corpus. *Prace Filologiczne* 63: 293-302.

von Waldenfels, Ruprecht & Meyer, Roland. 2006-. ParaSol, a Corpus of Slavic and Other Languages. <https://www.parasolcorpus.org/Ursynow/>

Weiss, Daniel. 2004. The rise of an indefinite article: The case of Macedonian *eden*. In Bisang, Walter, Himmelmann, Nikolaus & Wiemer, Björn (eds.) *What makes Grammaticalization? A look from its fringes and its components*, 139-165. Berlin/New York: Mouton de Gruyter.
<https://doi.org/10.1515/9783110197440.2.139>

Appendix. The number of occurrences of *'one' + N*

Table 1. All languages (subcorpus data combined)

	RUS	UKR	MKD	BUL	CES	HBS
numeral	208	201	191	146	125	142
derived/idiomatic	81	91	62	19	57	21
one+other	47	49	23	66	75	37
specific/presentative	42	42	114	138	86	88
non-spec/gen/pred	0	0	10	31	1	0
Total	378	383	400	400	344	288

Table 2. Russian (subcorpus data)

	Drina	Master	Lehkost	Potter	All
numeral	45	64	51	48	208
derived/idiomatic	20	19	22	20	81
one+other	21	6	16	4	47
specific/presentative	14	11	11	6	42
Total					378

Table 3. Ukrainian (subcorpus data)

	Drina	Master	Lehkost	Potter	All
numeral	47	60	36	58	201
derived/idiomatic	21	17	34	19	91
one+other	25	8	13	3	49
specific/presentative	7	15	10	10	42
Total					383

Table 4. Macedonian (subcorpus data)

	Drina	Master	Lehkost	Potter	All
numeral	43	56	45	46	190
derived/idiomatic	19	13	26	5	63
one+other	9	5	7	2	23
specific/presentative	26	22	21	44	114
non-spec/gen/pred	3	4	1	3	10
Total	100	100	100	100	400

Table 5. Bulgarian (subcorpus data)

	Drina	Master	Lehkost	Potter	All
numeral	37	39	29	41	146
derived/idiomatic	20	15	28	3	66
one+other	7	5	5	2	19
specific/presentative	32	37	28	41	138
non-spec/gen/pred	4	4	10	13	31
Total	100	100	100	100	400

Table 6. Czech (subcorpus data)

	Drina	Master	Lehkost	Potter	All
numeral	22	39	32	32	125
derived/idiomatic	20	13	34	8	75
one+other	26	12	12	7	57
specific/presentative	20	32	16	18	86
non-spec/gen/pred				1	1
Total	88	96	94	66	344

Table 7. Serbian (subcorpus data)

	Drina	Master	Potter	All
numeral	31	57	54	142
derived/idiomatic	20	10	7	37
one+other	14	5	2	21
specific/presentative	35	28	25	88
Total	100	100	88	288