

Krzysztof Hwaszcz, Hanna Kędzierska

Institute of English Studies, University of Wrocław

The Rise of an Indefinite Article in Polish: An Appraisal of Its Grammaticalisation Stage (Part 1)

Abstract

The aim of this paper is to assess the change of Polish numeral *jeden* ‘one’ into an indefinite marker in the view of the grammaticalization theory. Although Slavic languages are principally believed not to possess articles, certain usages of *one* (e.g. in Bulgarian and Macedonian) demonstrate the same features as the ones ascribed to the usages of indefinite articles in non-Slavic languages, such as English, German or Italian. Language contact of article-possessing languages is often claimed to enhance the grammaticalisation process of an indefinite article (Heine and Kuteva 2006). This type of grammaticalisation is said to follow five distinctive stages: (i) numeral, (ii) presentative marker, (iii) specific marker, (iv) non-specific marker and (v) generalized article (e.g. Givón 1981, Heine 1997). We assessed that in the case of Polish, the grammaticalisation stage is that of a specific marker, with some occasional uses leaning towards the non-specific marker stage. The conclusion was supported by the results of 53 native speakers’ judgments as well as the diagnostic tests based on relevant literature.

Keywords

grammaticalisation, indefinite article, language change, language contact, numeral ‘one’, the Polish language

Streszczenie

Celem artykułu jest ocena etapu gramatykalizacji rodzajnika nieokreślonego w języku polskim. Choć przyjęło się uważać, że rodzajniki nie występują w językach słowiańskich, w niektórych z nich zaimki wywodzące się od liczebnika jeden (np. w języku bułgarskim czy macedońskim) nabyły część cech i funkcji rodzajników nieokreślonych, występujących w innych językach, takich jak angielski, niemiecki czy włoski. Proces gramatykalizacji rodzajnika nieokreślonego może zostać przyspieszony poprzez kontakt międzyjęzykowy (Heine i Kuteva 2006) i przebiega w pięciu etapach: (i) liczebnika, (ii) słowa wprowadzającego (ang. *presentative marker*), (iii) wyznacznika referencji szczegółowej, (iv) wyznacznika referencji nieokreślonej oraz (v) rodzajnika (np. Givón 1981, Heine 1997). Na podstawie przeprowadzonych badań (ocen 53 rodzimych użytkowników polskiego oraz testów diagnostycznych) można stwierdzić, że w przypadku języka polskiego proces gramatykalizacji osiągnął etap wyznacz-

nika referencji szczegółowej, okazjonalnie wykazując też cechy charakterystyczne dla dalszego etapu gramatyzacji (tj. wyznacznika referencji nieokreślonej).

Słowa kluczowe

proces gramatyzacji, rodzajnik nieokreślony, zmiany językowe, kontakt międzyjęzykowy, liczebnik ‘jeden’, język polski

1. Introduction¹

A number of studies devoted to the issue of grammaticalisation (e.g. Heine 1997; Bybee 2002; Heine and Kuteva 2006; van Gelderen 2011) have called attention to the way the same grammatical phenomena are expressed across languages. In this article, we wish to make an attempt at finding what triggers the grammaticalisation process and to show that cross-linguistic similarities are not coincidental. Instead, they could be viewed as the outcome of conceptualization processes, whose function is to minimize the linguistic effort (van Gelderen 2011). These processes are influenced by paralinguistic factors, such as geographical proximity and language contact.

One particular instance of a conceptualization process – upon which we concentrate in the present paper – is a means of encoding indefiniteness in languages. Apart from word order, adjectival and verbal agreement suffixes or case opposition, indefiniteness can be expressed via articles. It is the communication strategy of a given language to distinguish a particular entity from a group of entities. The use of indefinite articles is rooted in general human cognitive capacities – namely, individuating an entity against a common experience to develop a shared situational world. Although both indefinite and definite articles are used to express grammatical phenomena within the same cognitive area, there appears to be a lack of systematicity between the existence of indefinite and definite articles across languages. Interestingly, this lack of systematicity is also valid within the same language group. In this study, we focus on the emergence of an indefinite article in Polish, which belongs to the family of Slavic languages. With a cross-linguistic overview of the occurrence of articles in Slavic languages, we want to demonstrate that language contact plays a crucial role in the evolution of articles.

The aim of the quantitative study reported in this paper is to tackle the grammaticalisation process of the Polish article. The research was divided into two parts: the first one was a survey study that tested the acceptability rating for particular uses of the Polish indefiniteness marker *jeden* ‘one’ based on the answers

¹ We would like to thank Professor Bożena Rozwadowska and two anonymous reviewers for their insightful and valuable comments on the paper as these comments led us to a considerable improvement of the work. This work was funded by a grant awarded to Krzysztof Hwaszcz by the Faculty of Letters, University of Wrocław (0420/2570/17).

given by 53 speakers of Polish (described in this part of the article); the second one was a corpus study in which over 20.000 sentences (dates ranged from 1992 to 2011) with the word *jeden* were analysed (discussed in Part 2 of the article).

2. What is grammaticalisation?

The last four decades have witnessed an increasing interest in grammaticalisation (Bybee 2002), the process by means of which a lexical item turns into a grammatical item. This mechanism triggers a change in the item's function and distribution (Heine and Reh 1984). One canonical example of grammaticalisation in English is the structure *to be going to*, as illustrated in (1).

- (1) a. movement: *I am going to school.*
 b. intention: *I am going to visit my parents.*
 c. future: *It is going to rain.* → *It's gonna rain.*

The literal interpretation in (1a) – which is most concrete and specific – in the course of time becomes more abstract and generalized (1b–1c). The raise of the new uses of a given construction boosts the number of appropriate contexts in which the construction can be used. *To be going to* was initially used only in its literal meaning (1a): in situations involving movement in which the subject was animate and volitional. Nowadays, it may be used even in such cases as in (1c), where no movement is to take place and the subject may be inanimate or extraposed. *Gonna* is the informal version of *to be going to* but as it becomes appropriate in a growing number of contexts, it appears more frequently in written English. Thus, the process of grammaticalisation is accompanied by the variation in both function and form.

The concept of the grammaticalisation path does not, however, presuppose a permanent time frame for the full cycle of evolution for a given linguistic phenomenon to take place; nor does it make the path obligatory in any way (Heine, Claudi and Hünemeyer 1991). The key principle of the path is to demonstrate the course the process follows once it has started and the potentiality for its expansion broken down into detectable stages. Consequently, the path is a reliable device for us to measure how far a particular form has progressed and anticipate what further developments or functions the form might receive.

2.1. The process of grammaticalisation of indefinite articles

Historical and comparative linguists accept that the indefinite articles emerge from the common source across languages: the numeral *one*. This grammaticalisation process follows five distinctive stages summarized in Table 1 below (adapted from Heine 1997).

Table 1. Indefinite article grammaticalisation

| Stage | Stage name | Function | Examples of languages |
|-------|---------------------|---|---------------------------------|
| | Numeral | numeral <i>one</i> exclusively | Swahili |
| 2. | presentative marker | a new referent is unknown to the hearer and known to the speaker; the new referent is mostly singular countable, specific and topical (i.e., intended to be used in the subsequent discourse) | Tarahumara |
| 3. | specific marker | a new referent is unknown to the hearer and known to the speaker; the new referent is mostly singular countable and specific, but not necessarily topical | Street Hebrew |
| 4. | non-specific marker | a new referent is unknown to the hearer and may be known or not to the speaker; the new referent is mostly singular countable, topical or not, specific or non-specific | English, German, Dutch, Punjabi |
| 5. | generalized article | a new referent is unknown to the hearer and may be known or not to the speaker; the new referent may be singular or plural countable or uncountable, topical or not, specific or non-specific | Spanish, Catalan, Portuguese |

The grammaticalisation process is a one-way path: once the article gains a new function, it cannot lose it and step back to the previous stage. At the same time, each new stage of grammaticalisation extends the range of functions fulfilled by the article, always retaining the previous ones.

Another implication inferred from the above description of functions is that the number of contexts in which the article can be found is gradually expanding, which in turn results in more relaxed conditions for the use of the article. Following the philosophy of semantic dilution over the grammaticalisation path, the lexical meaning of *one* keeps losing its lexical characteristics for the benefit of its grammatical meaning. The source meaning becomes bleached as grammaticalisation progresses. This change occurs with substantial phonetic erosion – fully developed indefinite articles frequently consist of fewer phonetic segments than the original numerals they arise from (Heine 1997).

The process of grammaticalisation could also be explained in terms of Economy Principles (Adger 2003, among many others) and more specifically, Feature Economy (van Gelderen 2011). Under this view, uninterpretable features are considered more economical than interpretable features because they can serve as probes and thus enable the syntactic derivation to proceed. The idea is based on the assumption that, within a Minimalist approach (Chomsky 1995, 2000, among others), *probes* (i.e., grammatical categories which possess uninterpretable features) search for *goals* (i.e., lexical categories which could

value these features via the Agree operation) (see (2a)). The search is limited to the goals which are c-commanded by the probes and it results in the deletion of the uninterpretable features (see (2b)).

- (2) a. [probe [goal]]
 [uF: ___] [iF: val]
 b. Agree ⇒ [probe [goal]]
 [~~uF: val~~] [iF: val]

(adapted from Zeijlstra 2012: 2)

To illustrate the point even further, in the derivation of a sentence *I read books*, little-*v* could be an example of a probe which has uninterpretable *phi*-features and searches for a noun or pronoun with interpretable *phi*-features (see (3)). The noun *books* can serve as its goal as it possesses a set of interpretable *phi*-features. Then, the probe and the goal undergo an Agree relation with each other (3a) and little-*v* can value its uninterpretable *phi*- features (3b) which then delete (3c).

(3)

PROBE-GOAL RELATION

- a. [_{VP} I [_{iφ}_[1sg] uCase_[_]] [_v √ [_{uφ}_[_]] [_{VP} READ BOOKS [_{iφ}_[pl] uCase_[_]]]]]]
 b. [_{VP} I [_{iφ}_[1sg] uCase_[_]] [_v √ [_{uφ}_[3pl]] [_{VP} READ BOOKS [_{iφ}_[pl] uCase_[_]]]]]]
 c. [_{VP} I [_{iφ}_[1sg] uCase_[_]] [_v √ [~~uφ~~_[3pl]] [_{VP} READ BOOKS [_{iφ}_[pl] uCase_[_]]]]]]

Now, let us compare the behaviour of numerals and pronouns. The number feature is interpretable on the first ones but uninterpretable on the latter ones. One could then imagine two possible scenarios with respect to such noun phrases as, for instance, *jedna książka* ‘one book’/‘a book’. If the word *jedna* is used as a numeral ‘one’ (4a), it cannot serve as a probe as it possesses an interpretable number feature. When the word *jedna* is used as a pronoun ‘a’, however, it needs to value its uninterpretable number feature, searches for a goal which possesses such a feature (in (4b), the word *książka* ‘a book’ serves as a goal) and thus makes the derivation proceed.

- (4) a. [_{NP} JEDNA [_{inumber}_[1sg]] KSIĄŻKA [_{iφ}_[sg,F] uCase_[_]]]]
 b. [_{NP} JEDNA [_{unumber}_[_]] KSIĄŻKA [_{iφ}_[sg,F] uCase_[_]]]]

Since children tend to interpret the language in most economical way, interpretable (i.e., less economic) features are reanalysed by them as uninterpretable (i.e., more economic) features. This would explain why numerals which possess interpretable number feature are reinterpreted as indefinite pronouns (which have an uninterpretable number feature). Later, the number feature becomes lost completely as the indefinite pronouns are further reanalysed as indefinite articles (see (5)).

- (5) a. numeral > indefinite pronoun > indefinite article
- b. i-F > u-F > zero

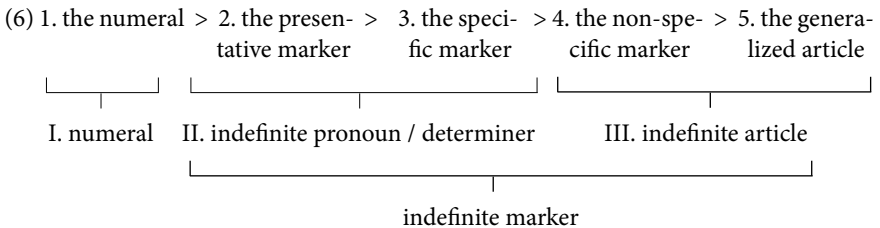
(on the basis of: van Gelderen 2011: 102)

To recapitulate, the grammaticalisation process starts from the function of quantification, through highly limited use of the incipient article, which initially refers to specific, countable and singular nominal phrases, and with time extends its references to less specific, plural, uncountable and generic nominal phrases (Givón 1981: 50). From the point of view of language acquisition, this phenomenon may be explained in terms of Feature Economy (van Gelderen 2011). Since indefinite articles frequently bear excessive functional load, the indefinite marker can eventually give rise to two indefinite articles, which attain exclusive functions (Chung and Ladusaw 2003).

2.2. A cross-linguistic overview of grammaticalisation of indefinite articles in Slavic languages

Before proceeding to the investigation of different uses of the Polish numeral *jeden*, it is important to consider the point of the grammaticalisation process with respect to the definite and indefinite articles across Slavic languages, to which Polish belongs. Although it is widely accepted that Slavic languages do not possess indefinite articles as such, there has been an intense debate about the Slavic nominal domain – namely, the issue of indefiniteness and NP vs. DP status of languages (e.g. Progovac 1998; Bošković 2009a, 2009b; Pereltsvaig 2013).

Linguists tend to avoid the notion ‘indefinite article’ when they refer to the linguistic element functioning as a carrier of indefiniteness in a given Slavic language. Conversely, they tend to use the term ‘indefinite marker’ (Friedman 2002) or ‘incipient category’ (Heine and Kuteva 2006). To avoid confusion, let us present the terminology used in this paper for the different grammaticalisation stages of indefinite articles (6).



(adapted from Geist 2011)

This paper additionally employs the term ‘something like articles’ (in the literature, used interchangeably with the term ‘article-like determiners’) to express that the numeral ‘one’ – in its original form – still has not fully developed into

an indefinite pronoun² but that some uses of it are congruous with the uses of indefinite markers in languages which have fully developed this category.

The structure of this section is based on the DP- versus NP-languages division, which is commonly postulated for Slavic languages (see, for instance, Bošković 2009a, 2009b). The distinction is formed in accordance with the assumption that certain properties of noun phrases could be associated only with article-less languages while others could be attributed only to languages which have articles. Importantly, indefinite articles are often claimed not to be situated in the DP (as in Bošković 2009a, 2009b). That is why, the NP/DP division is not expected to tell us much about the grammaticalisation stage of a numeral. However, it could be very informative with respect to the marking of indefiniteness and noun phrase structure in Slavic languages as the term NP languages refers to languages without articles or languages with an indefinite article only, while the term DP languages concerns languages with both indefinite and definite articles. Thus, we are going to distinguish between DP-languages (such as Bulgarian), NP-languages with something like articles (such as Polish) and languages with an indefinite article but no definite article (such as Slovenian or Upper Sorbian), which are considered NP rather than DP languages.

2.2.1. DP languages: Bulgarian and Macedonian

The most prominent Slavic language with respect to the presence of the indefinite article is Bulgarian. There is some discussion in the linguistic literature as to the nature of zero-marked noun phrases (henceforth NPs) and NPs with the head *edin* 'one'. Gorishneva (2013) argues for the semantic difference between these two phrases: bare NPs give priority to the interpretation of the corresponding referent, while NPs with the head *edin* are used to determine a new referent. From a pragmatic viewpoint, native speakers resort to their communicative strategy, upon which the choice between the two forms depends. The role of the referent is most crucial for this selection: when the language user regards the referent as salient, he or she tends to use *edin*-NPs; when the referent does not play an important contextual role, the speaker would use bare NPs. Consequently, non-salient NPs are viewed as semantically incorporated (cf. Farkas and de Swart 2003).

Thus, to formally describe the indefiniteness of NPs, it is crucial to take into consideration not only the distinctions between specific vs. non-specific or referential vs. non-referential, but also the token vs. type³ difference existing in

² In the article, the term pronoun is used in reference to items that function as nominal heads, and to items that are dependent upon nouns.

³ The distinction between type and its tokens is not restricted to linguistic corpora; rather, it is a useful metaphysical distinction which has a wide applicability in different areas, such as philosophy, science and linguistics (Wetzel 2009).

a given language (Gorishneva 2013). The term “token” indicates the overall number of words in a sentence or text, irrespectively of how many times each word is repeated. The term “type” indicates the number of specific (unrepeated) words in a sentence or text. This also holds true for the same discourse referents. That is, in the sentence “I needed to see a doctor but the doctor did not want to see me,” the actual number of token items is 15 but the number of type items is 10.

To make the type vs. token contrast clearer, let us consider some additional evidence. As far as German is concerned, this distinction can manifest itself even in the lexicon, which strongly suggests its relevance. There are two words in standard German – namely, *dasselbe* and *das gleiche* which mean ‘the same’.

(7) a. *Ich benutze das gleiche Auto wie mein Bruder.*

b. *Ich benutze dasselbe Auto wie mein Bruder.*

‘I use the same car as my brother.’

(adapted from Gorishneva 2013)

The pronoun *das gleiche* in (7a) implies that the car brand is the same but the speaker and his brother use different cars, whereas *dasselbe* in (7b) specifies that it is exactly the same car used by the speaker and his brother. Thus, *das gleiche* indicates type referentiality and *dasselbe* token referentiality. Although here it is a lexicalized – and therefore an overt – exemplification, it is most commonly covert in the early stages of the indefinite article grammaticalisation. The data from Bulgarian advocates the universal prediction put forward by Borthen (2003), which emphasizes the relevance of the token vs. type distinction: “In a language that has (something like) indefinite articles, lack of an indefinite determiner signals type-emphasis” (Borthen 2003: 226).

The notions of specificity and referentiality do not have clear-cut borders. These categories should rather be viewed as displaying gradience (cf. abstract referential NPs: Šmelev 2002). This signifies that the referential space may only be measured through the interplay of the categories with respect to their gradability (e.g. non-referential – weak referential – referential) due to the dynamic nature of the grammaticalisation process.

The interplay of pragmatic and semantic factors in Bulgarian regulates the use of *edin*-NPs and zero-marked NPs: *edin*-NPs, which are semantically scrutinized as token referents, are used to demonstrate discourse importance of the corresponding referent. On the other hand, zero-marked NPs, analysed as type referents, function as pragmatic markers of non-salience of the referent in the given discourse.⁴ The conclusion made for Bulgarian NPs indicates that the token vs. type distinction is greatly understudied and certainly requires further

⁴ For the rationale to keep this work concise, we decided not to include this lengthy discussion but rather to direct an inquiring reader to see Gorishneva (2013) for details and examples of Bulgarian NPs embedded in illustrative contexts.

analysis, which we wish to explore for Polish NPs (see Section 3.2. for a more detailed discussion of this issue).

Apart from indefinite articles, Bulgarian has also the definite article *ta* (for singular masculine) which takes a form of a phrasal suffix and is attached to its referent (Embick and Noyer 2001; Franks 2000). This definite article may appear on nouns, participles, numerals, adjectives or possessive pronouns. There is some controversy concerning the nature of the Bulgarian definite article because the article has the phonology of a suffix but the distribution of a clitic (Halpern 1995).

The Macedonian indefinite articles are comparable to the ones in Bulgarian (cf. Stojanov 1980). The indefinite article *eden* (formed from the numeral *eden* 'one') is mostly used in colloquial natural speech, while in formal speech or the written form, its use tends to be more restricted – there is a tendency to exclude the uses of *eden* as an article. Nonetheless, in the grammaticalisation process it, without any doubt, occupies the place of the specific marker (the third stage) because it may occur with specific abstract nouns, whether they are countable or uncountable, but appears mostly in the singular form (Weiss 2004). The readings of *eden* are parallel to the Polish *jeden* with the respect that both can function as adjectives (with the interpretation 'same, equal' or 'unique').⁵ From a geographical perspective, languages that occur on the Southern Balkan peninsula follow virtually the same grammaticalisation path of the indefinite article (Greek, Bulgarian, Macedonian and Albanian: in the order of the most advanced to the least advanced), at least when it comes to the referential statuses (Weiss 2004: 157). However, there exist certain differences between the uses of indefinite articles even between Macedonian and Bulgarian: Friedman (2002) quotes examples of sentences from the novel *Baj Ganjo* which contain the article *edin* in Bulgarian version but the same sentences are substituted with noun phrases with a zero-marker of indefiniteness in the Macedonian translation.

The definite article *ta* in Macedonian, like in Bulgarian, takes a form of a phrasal clitic but the Macedonian definite article exhibits a three-way spatial organization (8), while the Bulgarian definite article simply has a postpositive function of definiteness marker (9).

- (8) a. *kniga-ta*
 book-the (that)
 b. *kniga-va*
 book-this (here)
 c. *kniga-na*
 book-that (there)
- (9) *kniga-ta*
 book-the

(adapted from Franks 2000)

⁵ We discuss the grammatical function of *jeden* in Section 3.1.

Apart from the difference of marking the proximity in Macedonian and Bulgarian and in a slightly more restricted usage of the Macedonian *eden* over the Bulgarian *edin*, these two languages do not display any other significant differences and we will refer to them together as the only two Slavic languages which have both definite and indefinite articles.

2.2.2. NP languages with articles: Slovenian and Upper Sorbian

Slovenian is an interesting example because the numeral *en* ‘one’ is reported by Chierchia (1998) to have received new functions over the course of time and currently represents the final stage in the grammaticalisation process of indefinite articles. However, the use of this article is believed to be colloquial and only the speakers of Slovenian who come from the Littoral dialectal background use it on a daily basis (Runić and Juh 2017). Boškovič (2007) confirms Chierchia’s (1998) claim stating that – although spoken by a certain part of the speakers of Slovenian – this language has indefinite articles (but no definite ones) and, in his considerations, the indefinite does not belong to the DP.

Another Slavic language important for our considerations is Upper Sorbian, whose indefinite marker *jen* corresponds to the Modern English *a/an*. The numeral *jedyn* ‘one’ was the starting point for the development of the indefinite article in Upper Sorbian. Hence, *jen* is the only known instance across Slavic languages which has undergone phonetic erosion. This grammaticalisation change can be illustrated by the generalization formalized in (5a). Consequently, the syntactic alteration manifests itself in the loss of interpretable feature, as in (5b), where F is fulfilled by number.

The Upper Sorbian *jen* may be used in a generic sense (10a) or refer to a newly introduced referent, known neither to the speaker nor the hearer, as illustrated in (10b) below.

- | | | | | | | |
|---------|----------------------------|--------------|--------------|-------------|---------------|----------------|
| (10) a. | <i>Jen</i> | <i>tigor</i> | <i>jo</i> | <i>jene</i> | <i>wulke</i> | <i>zwěrjo</i> |
| | one | tiger.SG.NOM | BE.PRS.3SG | one | big | animal.SG.NOM. |
| | ‘A tiger is a big animal.’ | | | | | |
| | b. | <i>Ja</i> | <i>cem</i> | <i>jen</i> | <i>mikser</i> | <i>měč.</i> |
| | | 1SG.NOM | want.PRS.1SG | one | mixer.SG.ACC | have.INF |
| | ‘I want to have a mixer.’ | | | | | |

(Belaj and Matovac 2015)

Example (10a) is particularly interesting from the point of view of the token-type distinction discussed above (Section 2.2.1.): in this case, the indefinite marker *jen* denotes a referent which represents the type, or entire class. This is a very exceptional usage of ‘one’ in Slavic languages.

2.2.3. NP languages with ‘something like’ articles: Russian, Czech, Slovak and Croatian/Serbian

Russian is presented as one of the Slavic languages which has an incipient stage of indefinite article grammaticalisation (Heine and Kuteva 2006). It appears to be the case that the numeral *odin* ‘one’ may hardly be used as an indefinite marker except for contexts which involve either scientific discourse (11a) or – rarely – presentative marker (11b).

- (11) a. *ob odnom slučae palatalizacii*
of one case.SG.GEN palatalization.SG.GEN
‘of a (certain) palatalization phenomenon’
(Schroeder 2011)
- b. *Žyl da byl odin starik.*
live.PST.3SG.M PRT be.PST.3SG.M one old_man.SG.NOM
‘Once upon a time there was an old man.’
(Heine and Kuteva 2006)

Heine and Kuteva (2006) observe, however, that the use in (11b) as a presentative marker is acceptable only in certain contexts. A parallel use of *odin* in the example below (12) is very doubtful, with the observed difference in acceptability between (11b) and (12) most probably stemming from the fact that the referent in (12) is unique in a given context, there being only one tsar who can be easily identified by everyone.

- (12) *?Žyl da byl odin car.*
live.PST.3SG.M PRT be.PST.3SG.M one king.SG.NOM
‘Once upon a time there lived a king.’
(Aikhenvald and Dixon 2003)

The definiteness operator *jeden* in Czech is closely related to the Polish numeral *jeden* ‘one’ but its pragmatic load is somewhat different. Hlavsa (1975) observed that the use of this determiner as definiteness marker is optional and carries certain expressivity (13).

- (13) *Byl tu jeden člověk.*
be.PST.3SG here one person.SG.NOM
‘There was a man here.’

According to Bázlik (1991), Slovak *jeden* ‘one’ may carry the indefinite meaning of a noun phrase as well (13).

- (13) a. *Bolo tam jedno dieťa.*
‘There was a/one child.’

Poldauf (1969) specifies that the speaker uses the word *jeden* in (13) to signal that he or she does not wish be asked further questions about the referent (i.e., the non-salient use). The word *jeden*, which performs many other functions, is only occasionally used as the indefinite article and even in such exceptional uses, it is charged with an indicative attitude towards the addressee.

The last but not least on our list of languages with article-like determiners is Croatian/Serbian. In this language *jedan* 'one' may take the form of a numeral, adjective or an indefinite determiner. The article-like determiner in Croatian can even be used in predicative constructions (14a) or as a non-specific marker (14b).

(14) a. *Ti si jedna velika budala.*

'You are a (lit. one) big fool.'

(Belaj and Matovac 2015)

b. *Ivan želi oženiti jednu djevojku sa zelenim očima.*

'Ivan wants to marry a (lit. one) girl with green eyes.'

Another important characteristic of the article-like *jedan* is that, as opposed to the numeral *jedan*, it is not stressed (i.e., it acts as a proclitic, composing a unified accentual element with the subsequent word).

2.2.4. NP languages with no articles: Ukrainian and Belarusian

It is impossible for the numeral 'one' to function as either specific or non-specific indefinite in Ukrainian (15a)–(15b) and Belarusian (16a)–(16b).

(15) a. *Odnogo razu učitelja bilo bačeno*
 one time.SG.GEN teacher.SG.NOM was.PST.3SG seen.pass
v policejskij diljanci.
 in police office.SG.LOC
 'One day there was a teacher at the police office.'

b. *Ja xoču mati drugu.*
 I want.PRS.1SG have.INF friend.SG.ACC
 'I want to have a friend.'

(16) a. *Učorau večar pryexau sused.*
 yesterday in-the-evening came.PST.3SG neighbour.SG.NOM
 'The neighbour arrived last night'

b. *Kali laska, kupi mne gazet!*
 please buy.IMP me newspaper.SG.ACC
 'Please buy me a newspaper.'

(Heine and Kuteva 2006)

2.2.5. Indefiniteness in Slavic languages: a comparison

It is generally believed that Slavic languages do not possess articles – either definite or indefinite – as an independent grammatical category, with two

exceptions to this claim, i.e., Bulgarian and Macedonian definite articles in a form of a phrasal suffix. Nevertheless, the overview of the literature on the issue of (in)definiteness in Slavic languages (excluding Polish) compiled here made it apparent that certain usages of indefiniteness marker 'one' (as *edin*, *eden*, *en*, *jen*, *odin*, *jeden* or *jedan*) demonstrate the same features as the ones ascribed to the usages of indefinite articles in non-Slavic languages, such as English, German or Italian. Section 2.2 aimed at clearing up whether the Slavic indefinite determiners may be considered indefinite articles. We have reached the conclusion that the grammaticalisation process of indefinite articles has still not progressed towards the generalized article. Thus, it would be inappropriate to claim that Slavic languages reached the point at which indefinite articles constitute a separate grammatical category.

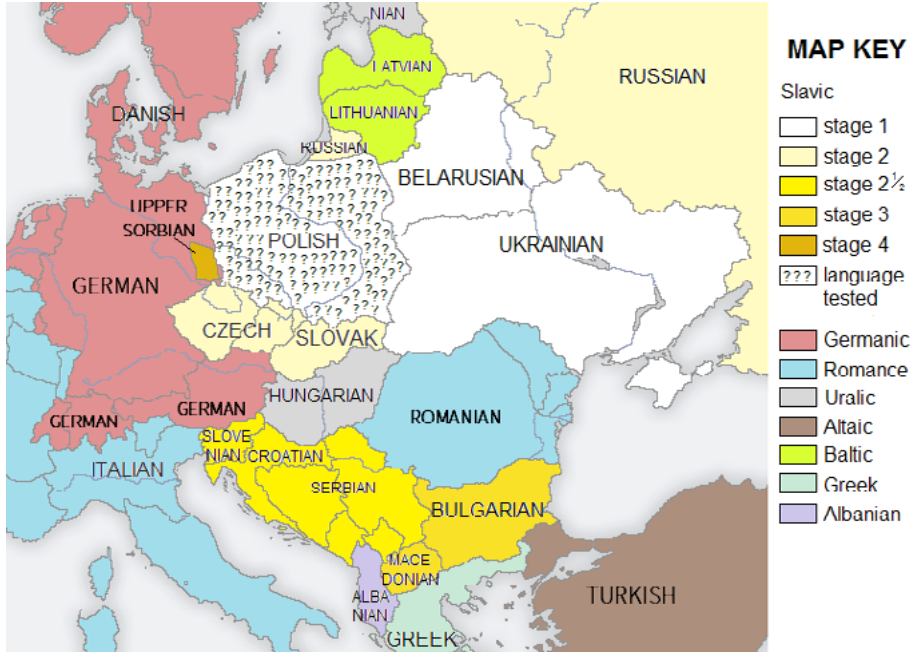
Table 2. Indefinite article grammaticalisation across Slavic languages (with an intermediate stage between a specific and non-specific marker)

| Stage name | Examples of Slavic languages | Indefinite marker name |
|------------------------------|------------------------------|-----------------------------|
| numeral | Ukrainian Belarusian | <i>odin</i> <i>adzin</i> |
| presentative marker | Czech/Slovak Russian | <i>jeden</i> <i>odin</i> |
| specific marker | Croatian Slovenian | <i>jedan</i> <i>en</i> |
| specific/non-specific marker | Bulgarian Macedonian | <i>edin</i> <i>edan</i> |
| non-specific marker | Upper Sorbian | <i>jen</i> |
| generalized article | - | - |

The collection of properties of the Slavic indefinite markers have been determined and discussed in this section. The capstone of this work is presented Table 2, where the languages are identified as to the stage where they presently are.⁶ The analysis of Map 1 can lead to the observation that there is indeed a geographical trend, which manifests itself in the following generalization: it seems that the closer a given Slavic language is to a Romance (Italian, Romanian), Germanic (German), Uralic (Hungarian), Greek (Greek) or Albanian (Albanian) language with indefinite articles, the more ahead in the development from the numeral 'one' to an indefiniteness marker this language is inclined to be. This conclusion seems to support the language contact theories, such as the Sprachbund Theory (Lyons 1999) which states that even genetically dispa-

⁶ The above analysis, although scrupulous, has been based on the examples collected from the literature. We have limited knowledge of most of the languages presented in the examination; thus, we apologize for any misstatements.

rate languages, yet geographically close ones, may affect each other. Hence, it is not surprising that the gradual geographical change progressed from article-possessing languages (mostly western and central Europe), through languages which have started this grammaticalisation process (the ones which have contact with western and central Europe), and finally to languages which do not seem to have any sign of indefinite articles (mostly eastern Europe).



Map 1. The stages of development of the indefinite article in Slavic languages (without Polish)
 Source: https://www.alternatehistory.com/wiki/lib/exe/detail.php?id=blank_map_directory%3Aall_of_europe&media=blank_map_directory:europa1.png (accessed: 19.07.2018).

One more issue which should be discussed at this point is the relation between the two accounts on the indefinite article grammaticalisation which were discussed in Section 2 (i.e., the strategy of substituting uninterpretable features for interpretable ones, described in Section 2.1. and the above-mentioned *Sprachbund* Theory). Although the first theory assumes that numeral-to-article grammaticalisation may occur spontaneously, it does not exclude the possibility of other factors becoming relevant for the process of language change: in fact, the observation that languages exert influence on one another which leads to faster extension of certain lexical items was also acknowledged by van Gelderen (2011: 26). According to her theory, the grammaticalisation process has its source in economy principles. However, it is not excluded that the change can later be facilitated by language contact and other sociolinguistic triggers postulated by the *Sprachbund* Theory. As the type of grammaticalisation discussed in the present

paper mostly affects spoken and uncodified language variety (at least at its initial stages), it seems to demand from the speakers a degree of bilingualism and/or a use of a mixed variety.⁷ However, with the increase in the number of international contact and in proficiency of the second received language (mostly English, Arabic, French, German, Spanish and Chinese; Beblavy et al. 2016: 5), it could be considered expected that the last three decades have witnessed an augmentation in the number of uses of indefinite articles in languages which commenced the process of indefinite article grammaticalisation.

To conclude, the reviewed properties of the Slavic indefinite markers undoubtedly prove that there are indefinite articles in Slavic languages (at least for Bulgarian, Macedonian and Upper Sorbian), but they have not developed a separate grammatical category. Thus, this phenomenon should not be considered a peripheral one.

3. Emergence of the indefinite article in Polish

The aim of this section is to establish the stage of the indefinite article grammaticalisation in Polish. In order to do that, a number of diagnostic tests were carried out on the basis of the tests which are most frequently conducted in the literature on indefinite articles (e.g. Topolińska 1981; Heine 1997; Grzegorzczkowska 2001; Heine and Kuteva 2006; Geist 2011; Runić and Juh 2017). Our account of indefinite article in Polish is based on the examples some of which were taken from the relevant literature, while others were verified by Polish native speakers with the use of online questionnaires. Assuming that the language contact scenario of indefinite article emergence (see Section 2) is right, Polish would be expected to have acquired at least stage 2 according to the classification proposed by Heine (1997).

3.1. Unmarked indefiniteness

Before the process of numeral grammaticalisation occurred in English, indefiniteness had been encoded in this language by means of word order. This is now the case in Polish. Szwedek (1973) provides the following example: while (17b) can follow the sentence in (17a), (17b) and (17c) cannot be interpreted as a sequence.

| | | | | |
|---------|-----------------------------------|---------------|--------------------|------------------|
| (17) a. | <i>W</i> | <i>pokoju</i> | <i>siedział</i> | <i>chłopiec.</i> |
| | in | room.SG.LOC | sit.PST.IPFV.3SG.M | boy.SG.NOM |
| | 'In the room, a boy was sitting.' | | | |

⁷ We would like to thank one of the anonymous reviewers for this remark.

| | |
|---------------------|---------------------|
| b. <i>Chłopiec</i> | <i>wyszedł.</i> |
| boy.SG.NOM | leave.PST.PFV.3SG.M |
| ‘The boy left.’ | |
| c. * <i>Wyszedł</i> | <i>chłopiec.</i> |
| leave.PST.PFV.3SG.M | boy.SG.NOM |
| ‘The boy left.’ | |

While the noun *chłopiec* ‘a boy’ in (17b) is interpreted as definite, the same noun in (17c) is indefinite. The example in (17a) demonstrates that definite nouns do not appear in the sentence final position in Polish, while indefinite nouns do not appear at the beginning of a sentence. It is, however, possible to emphasize the indefiniteness of the referent in (17a) by combining it with an indefinite pronoun such as *jakiś* or *jeden* (*W pokoju siedział jakiś/jeden chłopiec. Chłopiec wyszedł.* / ‘There was a boy sitting in the room. The boy left.’). The distinction between definite and indefinite entities in Polish is then expressed not only through the order in which noun phrases appear in a sentence but also through the use of adequate pronouns as well as prosodic cues.

3.2. Grammatical functions of the word *jeden*

The word *jeden* ‘one’ has a number of grammatical functions in the Polish language. First and foremost, it may be used as a **cardinal number** referring to the quantity (18).

| | | | | |
|---|------------------|-------------------|-----------------|-------------------------|
| (18) <i>Na</i> | <i>wykładzie</i> | <i>było</i> | <i>niewiele</i> | <i>studentów.</i> |
| at | lecture.SG.LOC | be.PST.3SG.N | FEW | student.PL.GEN |
| <i>Przyszła</i> | <i>jedna</i> | <i>dziewczyna</i> | <i>i</i> | <i>dwóch chłopaków.</i> |
| come.PAST.3SG.F | one | girl.SG.NOM | and | two boy.PL.GEN |
| ‘There were few students at the lecture. One girl and two boys came.’ | | | | |

Also, *jeden* may be used as an **adjective** used as a synonym for *samotny* ‘alone’ (19), *jednolity* ‘uniform’ (20), *identyczny* ‘identical’ (21)⁸, or used in order to emphasize the difference between the referent and some other entity or entities (22a–b) (Doroszewski 1958; Bańko 2000; Dubisz 2008). These additional functions of *jeden* are perhaps not surprising given the semantic relatedness between the numeral and adjectival uses, which are also present in other Slavic languages, such as Macedonian (Weiss 2004: 143).

⁸ The uses of *jeden* in examples (19)–(21) could perhaps be viewed as meta-predicative rather than qualitative expressions (cf. Danielewiczowa 2007) as they possess most of the qualities which were associated with meta-predicative expressions according to Danielewiczowa (2007): and more specifically, they cannot be used predicatively, are not stressed, cannot be modified by an adverbial, and are very unlikely to appear in apposition.

- (19) *Nikt nie był przygotowany. Jedna Maria*
 nobody NEG be.PST.3SG.M prepared one Mary.NOM
zrobiła zadanie domowe.
 do.PST.PFV.3SG.F homework.SG.ACC
 'Nobody was prepared. Only Mary did her homework.'
- (20) *Z kuchni dochodził jeden przeraźliwy*
 from kitchen.SG.LOC come.PST.IPFV3SG.M one appalling
hałas.
 noise.SG.NOM
 'A uniform, appalling noise was audible in the kitchen.'
- (21) *Mieszkaliśmy pod jednym dachem.*
 live.PST.1PL under one roof.SG.INS
 'We lived under the same roof.'
- (22) a. *Najpierw odwiedziła jedną koleżankę, a potem inną.*
 first visit.PST.PFV.3SG.F one friend.SG.ACC and then another
 'She visited one friend first, and then the other one.'
- b. *Przyszła jedna dziewczyna, potem druga.*
 come.PST.PFV.3SG.F one girl.SG.NOM then second
 'One girl came, and then the second one.'

Importantly, the use of *jeden* in (22b) has also been classified as an ordinal number (and not an adjective) in some of the dictionaries of Polish (e.g. Doroszewski 1958). Indeed, it seems possible to substitute the word *jeden* with *pierwszy* 'the first one,' in sentences such as (22b). However, the authors of more recent works (Bańko 2000, Dubisz 2008) seem to agree that such a function of *jeden* is rather to present the referent in opposition to some other entity (Dubisz 2008: 1280). Also, one should take into account that using the ordinal number *pierwszy* 'the first one' immediately implies the beginning of a series but the use of *jeden* involves no suggestion of that kind.

Interestingly, according to Doroszewski, *jeden* can also function as sort of **intensifier** usually following insulting noun phrases, such as: *zołza jedna* 'a shrew', *gamoń jeden* 'a berk', *urwis jeden* 'an imp', but also other noun phrases meant to be understood ironically. Thus, an expression such as *ekspert jeden* 'an expert' would imply that the referent is no expert at all.

More importantly for the purpose of our current study, however, *jeden* can also function as an indefinite pronoun, often used interchangeably with such pronouns as *jakiś* or *pewien* (23).

- (23) *Jeden chłopak dał mi tę*
 one boy.SG.NOM give.PST.3SG.M 1SG.DAT this.3SG.F.ACC
książkę.
 book.SG.ACC
 'A boy gave me this book.' / "One boy gave me this book.'

The sentence in (23) is, in fact, ambiguous. If embedded in an appropriate discourse context, it could be interpreted as an answer to the question “How many boys gave you this book?”. In such cases, the word *jeden* can usually be modified with a focus particle *tylko* ‘only’ (a test proposed in Topolińska 1981) or such modifiers as *dokładnie* ‘exactly’ or *wyłącznie* ‘strictly’. Much more frequently, however, sentences as 0 would be compatible with the use of *jeden* as an indefinite marker, which cannot be modified in the same ways as a numeral.

One more significant property of *jeden* proving that it can indeed be considered an indefinite marker is the fact that *jeden* can appear in the plural form (e.g. *jedni* ‘one.pl.masculine-human’ or *jedne* ‘one.pl.non-masculine-human’) (Saloni 1974a: 12; Saloni 1974b: 99–100). When *jeden* is used as numeral, such plural forms may appear only in the case of the so-called *pluralia tantum*, i.e., the nouns which do not have their singular forms, such as *jedne nóżyczki* ‘one.pl scissors’. Contrastingly, when used as an indefinite marker, *jeden* can combine with all countable nouns, resulting in such phrases as, for instance, *jedni mężczyźni* ‘one.pl.masculine-human men’ or *jedne psy* ‘one.pl.non-masculine-human dogs’.

3.3. *Jeden* as a presentative marker

The second stage of numeral grammaticalisation as described by Heine (1997) involves the use of a numeral as a presentative marker and, as we have already demonstrated in Section 2, this seems to be consistent with the data collected for many Slavic languages (see Table 2). Polish seems to be no exception with this respect as the word *jeden*, when used as an indefinite, is often used to introduce a new referent into the discourse (see (24)).

| | | | |
|---|----------------|----------------|-------------------------|
| (24) <i>Szczególnie</i> | <i>lubilem</i> | <i>jednego</i> | <i>koleżkę,</i> |
| particularly | like.PST.1SG.M | one | guy.SG.ACC |
| <i>lubilem</i> | <i>z nim</i> | | <i>rozmawiać (...).</i> |
| like.PST.1SG.M | with 3SG.M.INS | | talk.INF |
| ‘I particularly liked one guy, I liked talking to him.’ | | | |

(NKJP) (Janus and Przepiórkowski 2007)

One important factor which can influence the use of indefinites as presentative markers is the importance of the new referent in the discourse. As Farkas and de Swart (2003) and Geist (2011) point out, the pragmatic distinction between more and less relevant referents plays a crucial role in the use of Bulgarian *edin* as a presentative marker (see Section 2.1.1.2). This division also seems to be vital in the case of Polish. Thus, the use of the indefinite marker *jeden* seems to be more justified with contexts such as (25a) rather than (25b).

- (25) a. *Byłam akurat w mieście, więc odwiedziłam jedną koleżankę. Okazało się, że przeżyła niedawno ciekawą przygodę...* / 'I was in the city centre so I visited one friend. It turned out that she had recently had a very interesting adventure.'
- b. *Byłam akurat w mieście, więc odwiedziłam ?jedną koleżankę, poszłam też na pocztę i do parku...* / 'I was in the city centre so I visited ?one friend, went to the post office and to the park.'

3.4. Specific and non-specific uses of *jeden*: selected examples

When it functions as a pronoun, *jeden* has a tendency to be referentially used, similarly to such indefinite pronouns as *jakiś* or *pewien/pewny* (Topolińska 1981: 58). However, the use of the latter two pronouns usually involves clear predictions about their reference type. *Jakiś* is typically used when the speaker cannot identify the referent and *pewien* is more frequently encountered in the opposite contexts (i.e., the speaker being able to identify the referent). In order to verify which reference types can be denoted by *jeden*, we conducted an on-line questionnaire (see Appendix 1; for the rationale to keep the two parts relatively equal in size, we decided to include the Appendix in Part 2 of the present paper: Hwaszcz and Kędzierska, to appear).

According to Grzegorzczkowska (2001), the most frequent context in which this pronoun may be encountered involves such uses as (26).

- (26) *Jedna przyjaciółka mnie odwiedziła.*
 one friend.SG.F.NOM 1SG.ACC VISIT.PST.3SG.F
 'A friend visited me.'

The type of reference in (26) could be classified as *subjectively indefinite* reference (according to Grzegorzczkowska 2001). In other words, while the speaker knows who the referent is, it could be assumed that the addressee does not know it. As the world knowledge suggests, one is usually able to identify their friends, so the speaker in (26) should be able identify the referent. The reasons why he or she wants to present the referent as unidentified may be various. Most probably, being aware that the referent is unknown to the interlocutor, the speaker introduces the referent as unidentified also to him/herself. Such a use of an indefinite marker is associated with the third stage of grammaticalisation proposed by Heine (1996). Yet the conducted questionnaire revealed that only 18.9% of the respondents agreed with the interpretation of *jeden* mentioned above (i.e., only the subject being able to identify the referent) (see Table 3, Question 8). Also, only 22.6% of the subjects would use this pronoun while talking about a person that they know (see Table 3, Question 2). Such findings could suggest that examples like the one in (26) seem to be most frequent in colloquial speech. Interestingly, the pronoun *pewien* which is traditionally associated with such a context (e.g. Grzegorzczkowska 2001: 114),

was considered acceptable by a similar percentage of subjects (24.5%). Thus, it would seem to be dubious to conclude that the function of *jeden* as a specific marker is not yet well-established. Rather, though given many options, the interviewees simply seemed to favour definite pronouns such as *taki* in order to mark specific referents.

Table 3. The results of a questionnaire.

| Reference type questions ¹⁰ | Question ⁹ | Acceptability of a pronoun | | | |
|---|-----------------------|----------------------------|-------------|----------------|------------|
| | | Jeden | Jakiś | Pewien | Taki |
| unidentified to both the speaker and the addressee | 1 | 3 (5.7%) | 50 (94.3%) | 5 (9.4%) | 2 (3.8%) |
| unidentified to the addressee only | 2 | 12 (22.6%) | 6 (11.3%) | 13 (24.5%) | 19 (38.5%) |
| non-specific (generic) | 6 | 0 (0%) | 0 (0%) | 1 (1.9%) | 12 (22.6%) |
| Yes/no questions | | Yes | | No | |
| ironic/humorous explanation of <i>jeden</i> in (23) | 3 | 31 (58.5%) | | 22 (41.5%) | |
| overhearing explanation of <i>jeden</i> in (23) | 3 | 28 (52.8%) | | 25 (47.2%) | |
| existential use of <i>jeden</i> | 4 | 37 (69.8%) | | 16 (30.2%) | |
| optional use of <i>jeden</i> | 5 | 17 (32.1%) | | 36 (67.9%) | |
| Interpretation questions | | Only the speaker | Both | Neither | |
| a referent marked with <i>jeden</i> known to | 8 | 10 (18.9%) | 2 (3.8%) | 47 (88.7%) | |

⁹ See Appendix 1 (in Part 2 of the present paper: Hwaszcz and Kędzierska 2018, to appear).

¹⁰ The subjects could choose more than one correct answer. Also, it was possible to state that no answer was correct.

Further, Topolińska (1981) discusses other scenarios in which an unspecified argument may be used, depending on the speaker's and/or the addressee's ability to identify the referent. Some of them apply also to the pronoun *jeden*. First, one could imagine a scenario in which neither the speaker nor the addressee is able to identify the referent. In such cases, the pronoun *jeden* would rather be substituted with the pronoun *jakiś* (preferred by 94.3% of the respondents) rather than *pewien/pewny* (preferred by 9.4% of the respondents). 88.7% claimed that the use of *jeden* would be also compatible with a scenario in which the speaker cannot identify the referent (see Table 3, Question 8). This finding seems to be surprising in the light of the fact that only 5.7% of the respondents would use the pronoun themselves when referring to an unidentified entity (see Table 3, Question 1). This shows that such a use of *jeden*, though in vast majority acceptable, is far from being well-established in the language. Although the subjects, in general, would not use the word *jeden* while referring to an unidentified entity, they judged a sentence containing such a use as well-formed.

Also, it might be the case that both the speaker and the addressee are able to identify the referent. Such a scenario was considered convincing only by 3.8% of the respondents, though. Topolińska points out that the decision to present a referent as unidentified might stem from the speaker's fear of the conversation being overheard by someone else (cf. example (27)). Such an explanation of (27) was plausible to 52.8% of the respondents. Another possible explanation of the indefinite marker choice could be related to the sentence's aimed ironic or humorous overtone (which was considered a convincing explanation by 58.5% of the respondents).

(27) *Wczoraj odwiedziła mnie jedna nasza przyjaciółka.*
 yesterday visit.PST.IPFV3SG.F 1SG.ACC one our friend.SG.F.NOM
 'One of our friends visited me yesterday.'

Some other types of indefinite reference have been discussed by Grzegorzycykowa (2001). Most importantly, she distinguishes the cases in which an unspecified single referent is introduced in existential sentences (28) from cases such as (26) or (27). Also, she mentions situations in which the point of reference could be **optional**, i.e., associated with a specific but yet unidentified object (i.e., any entity belonging to a given set of entities) (compare examples (29) and (29)).

(28) a. *Jakiś człowiek bierze teraz ślub.*
 some man.SG.NOM take.PRS.3SG now wedding.SG.ACC
 'Some man is getting married now.' (= 'There exists a man who is getting married now').
 b. *Jeden człowiek bierze teraz ślub.*
 one man.SG.NOM take.PRS.3SG now wedding.SG.ACC
 One man is getting married now.' (≠ 'There exists a man who is getting married now').

- (29) a. *Podaj mi jakikolwiek długopis.*
 give.IMP 1SG.DAT any_sort_of pen.SG.ACC
 ‘Give me a pen.’
- b. *Podaj mi jeden długopis.*
 give.IMP 1SG.DAT one pen.SG.ACC
 ‘Give me one pen.’
- c. *Podaj mi jakiś długopis.*
 give.IMP 1SG.DAT some pen.SG.ACC
 ‘Give me some pen.’

The sentence in (28) could imply that the speaker is able to identify the referent, thus making the situation parallel to the one described in (26). Interestingly, however, the majority (i.e., 66.9%) of the respondents considered the *existential* interpretation of *jeden* in sentences like (28) correct. The situation described in (29) is less straightforward. On the one hand, the word *jeden* in (29) can be substituted with the phrase *tylko jeden* ‘only one’ and, consequently, it could perhaps be viewed as a numeral (which was the interpretation of 88.9% of the respondents). Under another sets of circumstances, it could, however, be substituted with the indefinite pronoun *jakiś* which makes it likely to be interpreted in a similar way as (29). Such an interpretation was considered acceptable by 32.1% of the respondents. Again, the interpretation here seems to depend on a wider discourse or sentential context.

To conclude, the pronoun *jeden* is considered acceptable when used as a specific marker only by some native speakers of Polish (22.6% of the respondents declared that they would use it themselves). *Jeden* could rarely be used

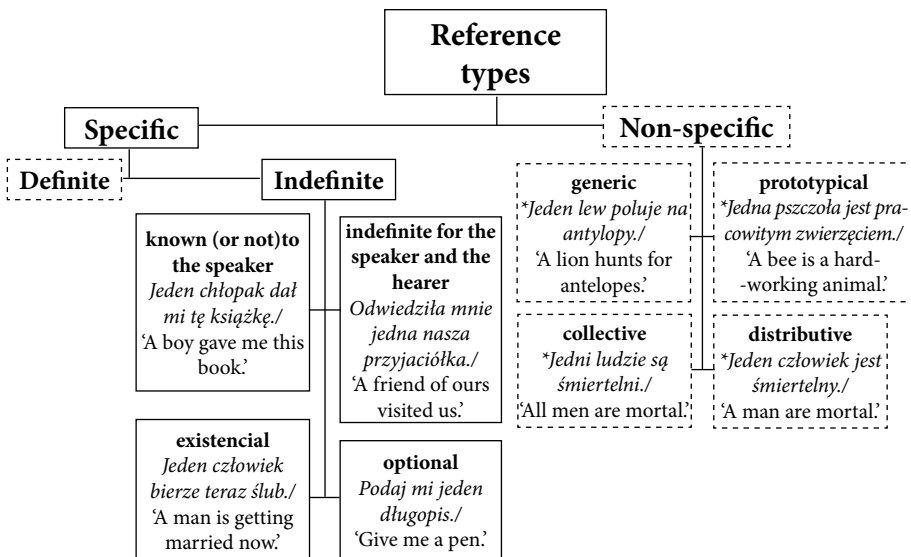


Figure 1: *Jeden* referentially-used: possibilities

if the speaker is not able to identify the referent (which would be compatible with the Stage 4 of grammaticalisation). Another indefinite pronoun (and more specifically, *jakiś*) is preferred in such contexts. Interestingly, when encountered out of context, *jeden* would most likely be interpreted as a reference to an entity which is unidentified by both the speaker and the addressee. This confirms that the five stages of grammaticalisation proposed by Heine et al. (1991) cannot be treated as separate and may, in fact, overlap one another (Lubańska 2009: 11). In the case of Polish, it seems that the use of an indefinite marker which is characteristic for Stage 4 is considered acceptable when processed but would be produced reluctantly. When it comes to the distinction between specific and non-specific reference, *jeden* is only acceptable in Polish when combined with specific indefinite referents (see Figure 1; based on Grzegorzyczkowa 2001: 112).

3.5. *Jeden* as an indefinite article: diagnostic tests

As it was demonstrated in the previous section, the Polish indefinite marker *jeden* can be used as a specific (but not as a non-specific) marker. This would suggest that the numeral grammaticalisation in Polish has already reached at least stage three on Heine et al.'s (1991) scale. Also, due to the fact that the use of an indefinite marker is on no account obligatory in Polish¹¹, it seems obvious that the grammaticalisation process cannot be considered complete. However, in order to fully determine the development of the process of numeral grammaticalisation, it is necessary to perform some tests concerning the last stage of the process (i.e., the emergence of an indefinite article) on Polish material.

First, a fully-grammaticalised indefinite marker would be expected to take both wide and narrow scope reading with intentional operators (Abush 1994, Runić and Juh 2017). It is a well-known fact that English sentences such as: *Kate wants to marry a prince* are ambiguous between their wide-scope readings (Kate wants to marry a **particular** prince) and narrow-scope reading (Kate wants to marry **any** prince). In Polish, only wide scope reading is possible when it comes to noun phrases with *jeden*, such as (30). Then, only the ending in (31) and not in (31) would be compatible with the sentence in (30).

| | | | | |
|----------------------------------|----------------|-----------------|----------------|---------------------|
| (30) <i>Kasia</i> | <i>pragnie</i> | <i>poślubić</i> | <i>jednego</i> | <i>hydraulika</i> . |
| Kate.NOM | want.PRS.3SG | marry.INF | one | plumber.SG.ACC |
| 'Kate wants to marry a plumber.' | | | | |

¹¹ In some languages with something-like articles, NPs which undergo topicalization require the occurrence of indefinite markers (Geist 2011, p. 137). It is similar in Polish as *jeden* frequently occurs in a topic position. However, it can be substituted with other indefinite markers such as *pewien* (as in *Jedna/pewna kobieta spotkała pewnego razu...* / 'One/some woman met once...' but not: **Kobieta spotkała pewnego razu...* / '*Woman met once...').

- (31) a. *Już nam go przedstawiła.*
 already 1PL.DAT 3SG.GEN introduce.PST.PFV.3SG.F
 'She has already introduced him to us.'
- b. *Jeszcze żadnego nie spotkała.*
 yet none NEG meet.PST.PFV.3SG.F
 'She has not met any yet.'

Similarly, indefinites marked by *jeden* cannot escape syntactic islands formed by a relative clause (a test proposed by Geist 2011). *Jeden* in (32) can only take wide scope, i.e., it must denote one particular professor while an English equivalent of the sentence (*Kate read every book that a teacher has recommended to her*) is again ambiguous between the narrow- and wide-scope reading.

- (32) *Kasia przeczytała wszystkie książki, które*
 Kate.NOM read.PST.PFV.3SG.F all book.PL.ACC that
polecił jej jeden nauczyciel.
 recommend.PST.PFV.3SG.M 3SG.F.DAT one teacher.SG.M.NOM
 'Kate read every book that a teacher has recommended to her.'

Another test proposed, among many others, by Geist (2011) involves the use of an indefinite functioning both predicatively and generically (with the two stages being often considered equal). When it comes to the first use, NPs denoting a profession or social status do not combine with *jeden* (**Jan jest jednym lekarzem* / '*John is one doctor') and thus cannot be used predicatively. Interestingly, however, Bulgarian *edin* in an intensifying function (very similar to such phrases as *głupek jeden* / 'a jerk' or *ekspert jeden* / (ironically) 'an expert', mentioned previously in Section 3.1) is considered to be a kind of predicative use by Geist (2011: 140).

When it comes to the question whether *jeden* can be used generically, the answer would be even less dubious than in the case of the predicative use. Polish equivalents of such sentences as: *A woman is always right* cannot contain indefinite markers (thus, they would be translated as: *Kobieta ma zawsze rację* rather than **Jedna kobieta ma zawsze rację*). The only possible exception to that rule would involve such sentences as: *Jedno jabłko kosztuje dziś (średnio) 80 groszy* 'Today, an apple costs (on average) 80 cents' or *?Jedno dziecko ma dziś dostęp do trzech urządzeń podłączonych do Internetu* 'Today, a child has access to three devices connected to the Internet'. Importantly, such a use of *jeden* seems to be compatible only with sentences describing some kinds of quantitative measures, cf. *?Waga jednej mandarynki jest mniejsza niż jednej pomarańczy* 'The weight of a/one tangerine is smaller than that of an/one orange' vs. **Kolor jednej mandarynki jest ciemniejszy niż jednej pomarańczy* 'The colour of a/one tangerine is darker than that of an/one orange'. This would imply that even in a generic sense *jeden* functions as a numeral. It would be then difficult to determine whether such instances as the ones mentioned above

could indeed be considered a germ of a predicative use or could rather be interpreted as a mark of general cognitive resources (in this case, related to quantity perception) reflected in a language.

Moreover, *jeden* in Polish, similarly to *edin* in Bulgarian (Geist 2011: 144) cannot be interpreted in the scope of negation, which is yet another proof that it cannot be considered a fully-developed article. In other words, the sentence in (33) is not equivalent to English *He did not say a thing* unless the indefinite marker *jeden* is preceded by a particle *ani* (then, *Nie powiedział ani jednej rzeczy* would be treated as equivalent with *He did not say a thing*).

- (33) *Nie powiedział jednej rzeczy.*
 NEG say.PST.3SG.M one thing.SG.GEN
 'He didn't say one thing.'

Finally, an indefinite article is expected not to induce a scalar reading with universal quantifiers (e.g. Dayal 2004; Runić and Juh 2017). In a Slovenian sentence such as: *Vsak dom ima eno televizijo* 'Nowadays there is a TV set in every home', the indefinite would be understood as *at least one* rather than *exactly one* (Runić and Juh 2017). This seems to be also the case in Polish: while comparing (34) with (35), *jeden* can be interpreted as *at least* in the first case, and in the latter one, it could be more likely substituted with the word *exactly*.

- (34) *Dziś w każdym domu jest jeden*
 today in every house.SG.LOC be.PRS.3SG one
televizor.
 TV_set.sg.nom
 'Today, there is one TV set in every house.'
- (35) *Dziś w każdym domu są dwa*
 today in every house.SG.LOC be.PRS.3PL two
televizory.
 TV_set.pl.nom
 'Today, there are two TV sets in every house.'

4. Estimating the stage of numeral grammaticalisation in Polish: final remarks

The cross-linguistic overview of grammaticalisation of indefinite articles in Slavic languages (Czech, Slovak, Russian, Croatian, Slovenian, Bulgarian, Macedonian and Upper Sorbian) as well as the contribution of new data for Polish have demonstrated that the indefiniteness marking in this language family has advanced further than a superficial examination could reveal. Additionally, all of the observations made in this paper are consistent with the unidirectionality principle and indicate that the numeral *one* changed into an indefinite marker

due to a language contact – not a single instance would suggest otherwise, whereby the indefinite article would change into the numeral *one* or would start losing the already acquired functions.

When it comes to the determination of the exact grammaticalisation stage of Polish *jeden*, we assessed that it is Stage 3, with some occasional uses leaning towards Stage 4. The conclusion was supported by the questionnaire results, the diagnostic tests as well as examples from the literature and the corpus. According to Heine (1991), for the numeral to serve as a specific marker, it has to meet the following requirements (and, in the case of Polish, it does): a new referent introduced by the marker should be unknown to the hearer and known to the speaker, the new referent should be mostly singular countable and specific, but not necessarily topical. All these requirements have been satisfied. Stage 4 (a non-specific marker) differs from Stage 3 in that the new referent could be unknown to both the speaker and the addressee (which has been attested in Polish but is not well-established), it may be topical (which is true for Polish because *jeden* can appear at the beginning of a sentence serving as a topic) and non-specific (as shown in Figure 1, its uses are not attested). In addition, based on the online questionnaire, the fourth stage of grammaticalisation seems to be only passive in the language, i.e. it was attested only in interpretation questions in which the testees were to assess who the referent is for the speaker and the listener. In this type of questions, the word *jeden* was associated with referents unknown to both the speaker and the addressee.

In Part 2 of our paper (Hwaszcz and Kędzierska, to appear), we will present the results of a corpus-based study whose aim was to measure the number of uses of *jeden* as an indefinite marker, and consequently gauge the directional tendency of the indefinite article in Polish and more accurately establish the stage of the process of grammaticalisation for the Polish indefinite marker.

References

- ABUSH Dorit (1994). The scope of indefinites. *Natural Language Semantics* 2(2), 83–135.
- ADGER David (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
- AIKHENVALD Alexandra Y., DIXON Robert Malcolm W. (eds.) (2003). *Studies in Evidentiality (Typological Studies in Language 54)*. Amsterdam and Philadelphia, PA: John Benjamins.
- BAŃKO Mirosław (2000). *Inny słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- BÁZLIK Miroslav (1991). Some notes on English and Slovak personal pronouns. *Brno Studies in English* 40, 25–34.
- BEBLAVY Miroslav, FABO Brian, LENAERTS Karolien (2016). The importance of foreign language skills in the labour markets of Central and Eastern Europe: an assessment based on data from online job portals. *CEPS Special Report* 129.

- 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.
- BORTHEN Kaja (2003). *Norwegian Bare Singulars*. Trondheim: Norwegian University of Science and Technology, PhD dissertation.
- BOŠKOVIČ Željko (2007). On the locality and motivation of move and agree: An even more minimal theory. *Linguistic Inquiry* 38(4), 589–644.
- BOŠKOVIČ Željko (2009a). More on the no-DP analysis of article-less languages. *Studia Linguistica* 63, 187–203.
- BOŠKOVIČ Željko (2009b). The NP/DP analysis of Slovenian. In *Proceedings of the University of Novi Sad Workshop on Generative Syntax 1*. Ljiljana Subotić (ed.), 53–73. Novi Sad: Filozofski fakultet u Novom Sadu.
- BYBEE Joan L. (2002). Cognitive processes in grammaticalisation. In *The New Psychology of Language* (Vol. II). Michael TOMASELLO (ed.), 145–167. New Jersey: Lawrence Erlbaum Associates Inc.
- CHIERCHIA Gennaro (1998). Plurality of mass nouns and the notion of ‘semantic parameter’. In *Events in Grammar*. Susan ROTHSTEIN (ed.), 53–103. Dordrecht: Kluwer Academic Publishers.
- CHOMSKY Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- CHOMSKY Noam (2000). Minimalist inquiries: the framework. In *Step by Step: Essays on Minimalist Syntax in Honour of Howard Lasnik*. Roger MARTIN, David MICHAELS and Juan URIAGEREKA (eds.), 89–115. Cambridge, MA: MIT Press.
- CHUNG Sandra, LADUSAW William (2003). *Restriction and saturation*. Cambridge, MA: The MIT Press.
- DANIELEWICZOWA Magdalena (2007). Przymiotniki nieprzymiotniki. O pewnym Niezwykłym typie wyrażen w języku polskim. *Zbornik Matice Srpske Za Slavistiku* 71–72, 223–236.
- DAYAL Veneeta (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27(4), 393–450.
- DOROSZEWSKI Witold (1958). *Słownik języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- DUBISZ Stanisław (2008). *Uniwersalny słownik języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- EMBICK David, NOYER Rolf (2001). Movement operations after syntax. *Linguistic Inquiry* 32(4), 555–595.
- FARKAS Donka F., DE SWART Henriëtte (2003). *The Semantics of Incorporation: From Argument Structure to Discourse Transparency*. Stanford: CSLI.
- FRANKS Steven (2000). The internal structure of Slavic NPs, with special reference to Bulgarian. In *Generative Linguistics in Poland 2 (Syntax and Morphology) (Proceedings of the GLIP-2 Conference)*. Adam PRZEPIÓRKOWSKI and Piotr BARAŃSKI (eds.), 53–70. Warsaw: Institute of Computer Science, Polish Academy of Sciences.
- FRIEDMAN Victor A. (2002). The Romani indefinite article in its historical and areal context. In *Was ich noch sagen wollte... A multilingual Festschrift for Norbert Boretzky on the Occasion of His 65th Birthday*. Birgit IGLA and Thomas STOLZ (eds.), 287–301. Berlin: Akademie Verlag.
- GEIST Ljudmila (2011). Bulgarian ‘edin’: The rise of an indefinite article. In *Formal Description of Slavic Languages: The Ninth Conference. Proceedings of FDSL 9*. Uwe

- JUNGHANNS, Dorothee FEHRMANN, Denisa LENERTOVÁ and Hagen PITSCH (eds.), 125–148. Frankfurt am Main: Peter Lang.
- GIVÓN Thomas (1981). On the development of the numeral 'one' as an indefinite marker. *Folia Linguistica Historica* 2(1), 35–53.
- GORISHNEVA Elena (2013). Bare vs. non-bare nouns: Two kinds of indefinites in Bulgarian. In *New perspectives on bare noun phrases in Romance and beyond*. Johannes KE-BATEK and Albert WALL (eds.), 301–328. Amsterdam: John Benjamins Publishing.
- GRZEGORCZYKOWA Renata (2001). *Wprowadzenie do semantyki językoznawczej*. Warszawa: Wydawnictwo Naukowe PWN.
- HALPERN Aaron (1995). *On the Placement and Morphology of Clitics*. Stanford: CSLI Publications.
- HEINE Bernd (1997). *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
- HEINE Bernd, CLAUDI Ulrike, HÜNNEMEYER Friederike (1991). *Grammaticalisation: A Conceptual Framework*. Chicago: University of Chicago Press.
- HEINE Bernd, KUTEVA Tania (2006). *The Changing Languages of Europe*. Oxford: Oxford University Press.
- HEINE Bernd, MECHTILD Reh (1984). *Grammaticalisation and Reanalysis in African Languages*. Hamburg: Helmut Buske Verlag.
- HLAVSA Zdeněk (1975). *Denotace objektu a její prostředky v současné češtině*. Prague: Academia.
- HWASZCZ Krzysztof, KĘDZIERSKA Hanna (to appear). The rise of an indefinite article in Polish: An appraisal of its grammaticalisation stage (Part 2). *Studies in Polish Linguistics* 13(3).
- JANUS Daniel, PRZEPIÓRKOWSKI Adam (2007). Poliqarp 1.0: Some technical aspects of a linguistic search engine for large corpora. In *The proceedings of Practical Applications in Language and Computers PALC 2005*. Jacek WALIŃSKI, Krzysztof KREDENS and Stanisław GÓDŹŹ-ROSKOWSKI (eds.). Frankfurt-am-Main: Peter Lang.
- LUBAŃSKA Maja (2009). On the emergence of the indefinite article in Polish. In *Pase Papers 2008. Studies in Language and Methodology of Teaching Foreign Languages*. Marek KUŹNIAK and Bożena ROZWADOWSKA (eds.), 63–78. Wrocław, Oficyna Wydawnicza ATUT.
- LYONS Christopher (1999). *Definiteness*. Cambridge: Cambridge University Press.
- PERELTSVAIG Asya (2013). Noun phrase structure in article-less Slavic languages: DP or not DP? *Language and Linguistic Compass* 7(3), 201–219.
- POLDAUF Ivan (1969). The so-called medio-passive in English. *Prague Studies in English* 13, 15–34.
- PROGOVAC Ljiljana (1998). Determiner phrase in a language without determiners. *Journal of Linguistics* 34(1), 165–179.
- RUNIĆ Marija, JUH Matej (2017). *The numeral 'en' revisited*. Poster session presented at the conference Formal Description of Slavic Languages 12.5. Slovenia: Nova Gorica, December 9, 2017.
- SALONI Zygmunt (1974a). Klasyfikacja gramatyczna leksemów polskich. *Język polski* 54(1), 3–13.
- SALONI Zygmunt (1974b). Klasyfikacja gramatyczna leksemów polskich. *Język polski* 54(2), 93–101.

- SCHROEDER Christoph (2011). Articles and article systems in some areas of Europe. In *Pragmatic Organization of Discourse in the Languages of Europe. Vol. 8 of Empirical Approaches to Language Typology*. Giuliano BERNINI and Marcia L. SCHWARTZ (eds.), 545–611. Berlin and Boston: Mouton de Gruyter.
- STOJANOV Stojan (1980). *Gramatičeskata kategorija opredelenost v bālgarski ezik (členuvane na imenata)*. Sofija: Narodna prosveta.
- ŠMELEV Aleksej (2002). *Russkij jazyk i vnezjazykovaja dejstvitel'nost'*. Moscow: Jazyki slavjanskoj kul'tury.
- SZWEDEK Aleksander (1973). A note on the relation between the article in English and word order in Polish. *Papers and Studies in Contrastive Linguistics* 2, 213–220.
- TOPOLIŃSKA Zuzanna (1981). *Remarks on the Slavic noun phrase*. Wrocław: Zakład Narodowy im. Ossolińskich Wydawnictwo Polskiej Akademii Nauk.
- VAN GELDEREN Elly (2011). *The Linguistic Cycle: Language Change and the Language Faculty*. New York: Oxford University Press.
- WEISS Daniel (2004). The rise of an indefinite article: The case of Macedonian 'eden'. In *What makes grammaticalisation: A look from its fringes and its components*. Walter BISANG, Nikolaus P. HIMMELMANN and Björn WIEMER (eds.), 139–165. Berlin and Boston: Mouton de Gruyter.
- WETZEL Linda (2009). *Types and Tokens: On Abstract Objects*. Cambridge, MA: MIT Press.
- ZEIJLSTRA Hedde (2012). There is only one way to agree. *The Linguistic Review* 29(3), 491–539.

Krzysztof Hwaszcz
Hanna Kędzierska
Instytut Filologii Angielskiej
ul. Kuźnicza 22, 50-138 Wrocław
[krzysztof.hwaszcz(at)gmail.com]
[hannaewakedziarska(at)gmail.com]