Frequency of Use of Polish Numerals Does Not Influence Their Syntax

Jadwiga Linde-Usiekniewicz and Paweł Rutkowski*

1 Introduction

Numerals referring to the four lowest cardinalities differ morpho-syntactically from the rest of cardinals in Polish and many other Indo-European languages. In the present paper, we aim to analyze to what extent the above fact can be correlated with the frequency of use of the numerals in question. Heine 1997 proposes a cross-linguistic generalization, according to which lower numerals tend to be more adjectival than higher ones because they are used more frequently. We will confront this model with Polish frequency data (taken from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990) and argue that Heine's 1997 approach does not account for the fact that the class of Polish cardinal numerals is divided into three distinct subclasses.

2 Lower vs. Higher Numerals and Heine's 1997 Grammaticalization Model

As has been observed in the typological literature (cf. Hurford 2001), one of the most puzzling features of many numeral systems is the fact that cardinals are split into two classes. The lowest cardinals (this subset usually includes 'one', 'two', 'three' and possibly 'four') are often morpho-syntactically different from the rest of the lexical class in question. Hurford 2001 views this discontinuity at around 4 as a cross-linguistic phenomenon. By contrast, Hammarström 2004 argues that such a clear-cut split is typical of Indo-European only because it is inherited from Proto-Indo-European. Even if the lower/higher numeral distinction is not related to 4 in all language families, its special status in at least some natural languages seems intriguing.[1]

In its extreme form, the lower/higher numeral division manifests itself in the fact that numerals higher than 'three' or 'four' do not exist at all (higher magnitudes are referred to as 'many'). It is also often the case that the lowest numerals are less complex morphologically than higher ones (the latter are derived from the former by analogy to the arithmetic operations of addition or subtraction). In some languages, elements expressing low and high numerosities belong to different categories. Mithun's (1999) discussion of Wiyot (an Algic language of Northern California) provides an interesting example of such a situation. Wiyot uses verb stems *kuc-* 'be one', *dit-* 'be two', *dikh-* 'be three', *diyohw-* 'be four' to refer to the numbers 1 – 4 (directly inflected as verbs, or followed by a classifier, such as: *-okh* 'longish object', *-atk* 'roundish object', *-apł* 'hairlike object' etc.). However, higher numerosities are expressed with numerals (followed by the stem *hal-* 'so many' and a classifier):

(1) dikh-okh
be.three-longish
'three longish objects'

(2) takłaluk hal-okh
six so.many-longish
'six long objects'

The lower/higher numeral dichotomy may also influence certain inflectional properties of cardinals: the numerals in the range 1 – 4 are usually characterized by richer and more idiosyncratic declensions. Hurford 2001 gives several examples of this phenomenon. In some languages, only the numerals for 1 – 4 have more than one idiosyncratic inflectional (gender/case) form – see Tables 1 and 2.

The numerals in the range 1 – 4 tend to be syntactically subordinate to the quantified noun: for instance, they have to agree with the noun in terms of various inflectional features (typically, case and gender). On the other hand, higher numerals can often act as case-assigners. These two types of morpho-syntactic behavior are

* An earlier version of this paper was published in Polish as Linde-Usiekniewicz and Rutkowski 2003. We are indebted to Paweł Boguszewski, Magdalena Derwojedowa, Romuald Huszcza, Scott McClure, and Marek Świdziński for helpful comments on this analysis. The second author gratefully acknowledges the Junior Advanced Research Grant awarded by the Polish-American Fulbright Commission, which made it possible for him to spend the 2005-2006 academic year at Yale University.

[1] Rutkowski 2003 attempts to explain why 4 functions as a threshold by linking this fact to Cowan's 2001 "magical number four", i.e. to the observation that the storage capacity of human perception and short-term memory is limited (it seems to oscillate between 3 and 5 items).

illustrated in (3): in Latin, numerals such as *tres* 'three' agree with the case form of the quantified noun, whereas high numerals such as *centum* 'hundred' require the noun to assume genitive:

(3) tres centi puerorum
three-NOM hundreds-NOM boys-GEN
'three hundred boys'

Numeral	Greek	Icelandic	Welsh
'1'	3	3	1
'2'	1	3	2
'3'	2	3	2
'4'	2	3	2
'5'	1	1	1
'6'	1	1	1
'7'	1	1	1
'8'	1	1	1
'9'	1	1	1
'10'	1	1	1

Table 1. Number of idiosyncratic gender forms (Hurford 2001)

Numeral	Greek	Icelandic	Albanian
'1'	3	4	3
'2'	1	4	3
'3'	2	4	1
'4'	2	4	3
'5'	1	1	1
'6'	1	1	1
'7'	1	1	1
'8'	1	1	1
'9'	1	1	1
'10'	1	1	1

Table 2. Number of idiosyncratic case forms (Hurford 2001)

Polish is a very clear example of a language in which the distinction between the two syntactic case patterns shown in (3) correlates with the 4/5 split (the numeral for 4 agrees in case, whereas the numeral for 5 assigns genitive). The lowest four Polish cardinals (labeled A-numerals by Rutkowski 2001) behave like adjectives, i.e. they always agree with the quantified noun in case, number and gender, whereas cardinals such as *pięć* 'five' (Q-numerals) assign the so-called "Genitive of Quantification" to the following noun. However, Q-numerals act as case assigners in structural case contexts only (nominative or accusative). In the context of inherent cases (such as instrumental or dative), Q-numerals behave like adjectives and agree in case with the head noun. In the following examples (taken from Rutkowski 2006), the verb *lubić* 'like' requires accusative, whereas the verb *doradzać* 'advise' assigns dative:

(4) a. Cezary lubi trzy osoby. (A-numeral, structural context)
Cezary likes three-ACC people-ACC
'Cezary likes five people.'
(4) b. *Cezary lubi trzy osób.
Cezary likes three-ACC people-GEN

(5) a. Cezary doradza trzem osobom. (A-numeral, inherent context)
Cezary advises three-DAT people-DAT
'Cezary advises three people.'
(5) b. *Cezary doradza trzem osób.
Cezary advises three-DAT people-GEN

(6) a. Cezary lubi pięć osób. (Q-numeral, structural context)
Cezary likes five-ACC people-GEN
'Cezary likes five people.'
(6) b. *Cezary lubi pięć osoby.
Cezary likes five-ACC people-ACC

(7) a. Cezary doradza pięciu osobom. (Q-numeral, inherent context)
Cezary advises five-DAT people-DAT
'Cezary advises five people.'
(7) b. *Cezary doradza pięciu osób.
Cezary advises five-DAT people-GEN

Very high numerals such as *tysiąc* 'thousand' or *milion* 'million' (N-numerals) always assign genitive to the following noun. This means that, in terms of case assignment, they are fully nominal.

(8) a. Cezary lubi milion osób. (N-numeral, structural context)
Cezary likes million-ACC people-GEN
'Cezary likes one million people.'
(8) b. *Cezary lubi milion osoby.
Cezary likes million-ACC people-ACC

(9) a. Cezary doradza milionowi osób. (N-numeral, inherent context)
Cezary advises million-DAT people-GEN
'Cezary advises one million people.'
(9) b. *Cezary doradza milionowi osobom.
Cezary advises million-DAT people-DAT

The above pattern of case assignment is summarized in Table 3.

Genitive assignment	A-numerals	Q-numerals	N-numerals
in structural contexts	-	+	+
in inherent contexts	-	-	+

Table 3. Genitive assignment in Polish numeral expressions

Corbett 1978 analyses similar data from Russian and proposes the following two universals:

(10) The syntactic behavior of simple cardinal numerals will always fall between that of adjectives and nouns.

(11) If the simple cardinal numerals of a given language vary in their syntactic behavior, the numerals showing "nounier" behavior will denote higher numerals than those with less "nouny" behavior.

According to Heine 1997, the fact that some cardinals behave like nouns, whereas others resemble adjectives indicates that the adjectival ones have undergone a process of grammaticalization. This process involves a semantic shift: a conceptual transfer from a more concrete meaning to a more abstract one. The denotation of numerals which are subject to grammaticalization changes from real objects to abstract qualities (which are independent of the original object and can be applied to other objects). This transfer is parallel to the one underlying the development of other derived adjectives, such as color terms (a word for a fruit may be reanalyzed as an adjective referring to a particular color, e.g. 'orange').

If this analysis is on the right track, an explanation should be given to the fact that, in languages such as Polish, not all numerals have become adjectival, but only the lowest ones. Heine 1997 argues that the degree of grammaticalization ("adjectivalization") is positively correlated with the frequency of use of a given numeral. Therefore, the Polish pattern shown in Table 3 is predicted to be shaped by frequency factors: A-numerals must be used more often than Q-numerals, which in turn are more frequent than N-numerals. The above nominal-to-adjectival model seems to find *prima facie* confirmation in the diachronic development of Polish cardinals. As

discussed in Rutkowski 2002 and 2006, the Old Polish equivalents of present-day Q-numerals were not sensitive to their case environment: they always assigned genitive to the quantified noun (see Table 4).

Case context	Old Polish	Modern Polish	Gloss
Nominative	siedm grzechow seven-NOM sins-GEN	siedem grzechów seven-NOM/ACC[2] sins-GEN	'seven sins'
Accusative	sześć świadkow six-ACC witnesses-GEN	sześciu świadków six-ACC witnesses-GEN	'six witnesses'
Dative	sześcidziesiąt dział sixty-DAT cannons-GEN	sześćdziesięciu działom sixty-DAT cannons-DAT	'sixty cannons'
Instrumental	siedmią ran seven-INSTR wounds-GEN	siedmioma ranami seven-INSTR wounds-INSTR	'seven wounds'
Locative	ośmi lat eight-LOC years-GEN	ośmiu latach eight-LOC years-LOC	'eight years'

Table 4. Diachronic development of the syntax of Polish numerals

The change between Old and Modern Polish could be analyzed as an example of grammaticalization (in line with Heine's 1997 assumptions). The syntax of Q-numerals used to be identical with that of regular nouns (and N-numerals), however, it has become more "adjectival", due to conceptual transfer.[3] It should be noted that such an explanation would require us to assume that Old Polish Q-numerals were used more often than N-numerals (which have remained nominal until today). This Heine-type analysis of the syntactic development of Polish numerals could be represented in the following way ("<<<" stands for "are less nominal and used more frequently than"):

(12) A-numerals <<< Q-numerals <<< N-numerals

This analysis, although interesting as a general model, faces several problems. Heine's 1997 hypothesis assumes unidirectionality; in other words, he predicts that it is unlikely for an "adjectival" numeral to become more "nominal" as a result of grammaticalization. However, a kind of adjectival-to-nominal reanalysis seems to be taking place in Modern Polish. There are two variants of the numerals for 2 – 4, when used with masculine nouns in nominative (sentential subject) positions. Apart from the conservative adjectival forms *dwaj* 'two', *trzej* 'three' and *czterej* 'four', it is also possible to use the Q-numeral forms *dwóch* 'two', *trzech* 'three' and *czterech* 'four'. The latter do not agree in case with the quantified noun but, similarly to Q-numerals such as *pięć* 'five', act as genitive-assigners:

(13) Trzech mężczyzn spało.
three-NOM/ACC men-GEN slept
'Three men slept.'

This syntactic innovation contradicts Heine's 1997 proposal. A-numerals get reanalyzed as Q-numerals, which is the opposite of what he predicts (assuming Heine's 1997 line of reasoning, we would rather expect Q-numerals to have a tendency to become more adjectival as a result of further grammaticalization).[4]

The above piece of counterevidence is very specific, and it could be considered marginal from the point of view of Heine's 1997 model. However, one of the most crucial assumptions of his model, namely that the nominal-to-adjectival shift is driven by frequency factors, is very questionable as well. In the remaining part of this paper, we will confront Heine's 1997 proposal with the actual frequency of use of Polish cardinal numerals. We will argue that his generalization does not find empirical confirmation, and, therefore, it does not suffice to explain why the word for 5, as opposed to the word for 4, is not an A-numeral, and the word for 1000, as opposed to the word for 900, is not a Q-numeral.

[2] There is a controversy in the literature on Polish numeral expressions whether Q-numeral subjects are accusative or nominative (see Franks 1995, Przepiórkowski 1996, 2004, Rutkowski 2000, among others). We will not discuss this issue here because it does not directly influence our analysis.

[3] See Rutkowski 2002 and 2006 for an alternative analysis of this grammaticalization process, based on the proposal that the syntactic status of Modern Polish Q-numerals is functional, and not lexical.

[4] Note that no A-to-Q shift is possible when the numerals in the range 2 – 4 are used with feminine or neuter nouns (only A-type nominative forms of the numerals are available in these genders). Therefore, the class of A-numerals does not seem to be disappearing altogether (merging with Q-numerals).

3 Frequency Data from Polish

We have examined numeral data taken from Kurcz, Lewicki, Sambor, Szafran and Woronczak's 1990 frequency dictionary of Modern Polish. The dictionary is based on a corpus of 500,000 words, which was collected in 1963-1967. The texts included in the corpus were classified as belonging to one of the following registers/genres: popular science, short press reports, political commentary journalism, literary prose, drama. The dictionary also includes frequency data from spoken Polish, taken from Zgółkowa 1983.

From the point of view of Heine's 1997 hypothesis, the most interesting piece of data is the difference in frequency between the numeral *cztery* 'four' and *pięć* 'five'. Interestingly, as illustrated in Table 5 (which shows the total number of appearances of the numerals in the range 2 – 10 in the corpus[5]), the frequency of use of the numeral *pięć* 'five' is actually higher than the frequency of the numeral *cztery* 'four'.

Numeral	Frequency
dwa 'two'	936
trzy 'three'	568
cztery 'four'	373
pięć 'five'	431
sześć 'six'	240
siedem 'seven'	164
osiem 'eight'	221
dziewięć 'nine'	135
dziesięć 'ten'	202

Table 5. Total frequencies of the cardinals *dwa* 'two' – *dziesięć* 'ten'
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

This situation is far from unique cross-linguistically. Juillard and Chang-Rodriguez's 1964 frequency dictionary of Spanish shows that the numeral *cinco* 'five' has a higher frequency than the numeral *cuatro* 'four'. Hammarström 2004 points out that 'five' is slightly more frequent than 'four' in three Russian frequency dictionaries, whereas 'four' is slightly more frequent than 'five' in one French, one Latvian and one Italian frequency dictionary. It should be noted that, even if the word *pięć* 'five' were less frequent than *cztery* 'four' in Polish, Heine's 1997 proposal could be considered plausible only if the difference in frequency between the two numerals were significant, and not slight.

Moreover, it seems that all Polish cardinals derived from the base number 5 occur with a relatively high frequency. In each morphological series of numerals (*jeden* 'one' – *dziewięć* 'nine', *jedenaście* 'eleven' – *dziewiętnaście* 'nineteen', *dziesięć* 'ten' – *dziewięćdziesiąt* 'ninety' and *sto* 'one hundred' – *dziewięcset* 'nine hundred'), the derivatives of *pięć* 'five' (such as *piętnaście* 'fifteen') are more frequent than the derivatives of *cztery* 'four' (such as *czternaście* 'fourteen') – see Table 6. We take the above data to illustrate the influence of extralinguistic (arithmetic) factors on the frequency of cardinals. The fact that numerals related to 5 occur very often is clearly caused by the base status of this number in the decimal system (it is used in approximations etc.). Interestingly, the special status of the numeral *pięć* 'five' is less salient in the frequency data taken from more colloquial varieties of Polish.

Numeral	Frequency	Numeral	Frequency	Numeral	Frequency
czternaście 'fourteen'	46	czterdzieści 'forty'	185	czterysta 'four hundred'	65
piętnaście 'fifteen'	109	pięćdziesiąt 'fifty'	262	pięćset 'five hundred'	102
szesnaście 'sixteen'	39	sześćdziesiąt 'sixty'	229	sześćset 'six hundred'	67

Table 6. Total frequencies of cardinals derived from the numerals *cztery* 'four', *pięć* 'five' and *sześć* 'six'
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

[5] The word *jeden* 'one' is not included in Table 5 since, apart from being a numeral, it also functions as an indefinite pronoun ('some'). Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 do not give the frequencies for these two uses separately.

As shown in Table 7, the relative frequency of the word *pięć* 'five' is much higher in scientific texts, and press reports or commentaries than in drama and spoken language. The latter two include less numeric data than formal texts, so the "approximation" effect is alleviated. The same is true for Polish child language – see Table 8. It should be noted, however, that even in colloquial or child Polish, A-numerals are not significantly more frequent than the numeral *pięć* 'five'. Therefore, there is no reason to suggest that the more "nouny" character of the latter is related to how frequently it is used.

Numeral	Popular science	Short press reports	Political commentary journalism	Literary prose	Drama	Spoken language
dwa 'two'	199	322	157	147	111	218
trzy 'three'	118	240	82	58	70	84
cztery 'four'	65	135	67	33	23	65
pięć 'five'	83	187	97	25	39	57
sześć 'six'	34	129	52	13	12	19

Table 7. Total frequencies of the cardinals *dwa* 'two' – *sześć* 'six' in different styles (data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 and Zgółkowa 1983)

Numeral	Frequency
dwa 'two'	207
trzy 'three'	90
cztery 'four'	47
pięć 'five'	50

Table 8. Total frequencies of the cardinals *dwa* 'two' – *pięć* 'five' in child language (data from Zgółkowa and Bułczyńska 1987)

According to Hammarström 2004, Heine's 1997 frequency analysis is on the right track, but it needs to be combined with the assumption that base numbers are "cognitive reference points" (Rosch 1975, Sigurd 1988), and therefore, they are used more often. However, such an amendment to Heine's 1997 theory makes it lose its explanatory power. If we argue that, being a cognitive reference point, the numeral for 5 is likely to be used more frequently than the numeral for 4, we cannot at the same time claim that the numeral for 4 is less "nouny" than 5 due to a negative correlation between frequency and "nouniness". The two claims are simply contradictory.

We argue that, in the light of the above data, Heine's 1997 model does not seem to explain the lower/higher numeral split in Polish. In order to be compatible with the Modern Polish data, the frequency-based analysis would have to imply that there was a stage in the development of Polish numerical constructions at which the frequency of use of the four lowest numerals was much higher than that of all the other numerals, including base numbers (or Hammarström's 2004 "cognitive reference points"). In our opinion, it is much more plausible to assume the opposite: because of its special status in arithmetic, the numeral referring to 5 must have always occurred more often than the numeral referring to 4, at least in the languages of those societies which were using the decimal system.[6]

Other data also seem to contradict Heine's 1997 hypothesis. As shown in Table 9, the numerals *tysiąc* 'one thousand' and *dziewięćset* 'nine hundred' are among the most frequent cardinals. The corpus used by Kurcz, Lewicki, Sambor, Szafran and Woronczak was collected in the 1960s. Therefore, the surprisingly high frequency of *tysiąc* 'one thousand' and *dziewięćset* 'nine hundred' can be explained away by the fact that the two numerals often appear in dates (see (14)). However, it is impossible to claim that the nominal properties of the N-numeral *tysiąc* 'one thousand' are related in any way to low frequency.

[6] Therefore, Rutkowski 2003 suggests that the 4/5 threshold results from the fact that the four lowest numerals appeared in the lexicon of Indo-European earlier than the decimal system was developed. According to this analysis, the present-day lower/higher numeral dichotomy is related to different origins of the two numeral classes in question, and not to frequency factors: the numerals for 1 – 4 developed earlier and independently from the other counting words.

(14) rok tysiąc dziewięćset dziewięćdziesiąty
year thousand nine-hundred ninetieth
'the year nineteen ninety'

Numeral	Frequency
tysiąc 'thousand'	1207
dwa 'two'	936
dziewięćset 'nine hundred'	572
trzy 'three'	568
pięć 'five'	431
dwadzieścia 'twenty'	400
cztery 'four'	373
sto 'one hundred'	330
trzydzieści 'thirty'	266
pięćdziesiąt 'fifty'	262
sześć 'six'	240
sześćdziesiąt 'sixty'	229
osiem 'eight'	221
dziesięć 'ten'	202
czterdzieści 'forty'	185
siedem 'seven'	164
dwieście 'twenty'	161
dziewięć 'nine'	135
osiemdziesiąt 'eighty'	124
siedemdziesiąt 'seventy'	123
piętnaście 'fifteen'	109
pięćset 'five hundred'	102
trzysta 'three hundred'	101
dwanaście 'twelve'	87
osiemset 'eight hundred'	85
dziewięćdziesiąt 'ninety'	79
sześćset 'six hundred'	67
czterysta 'four hundred'	65
siedemset 'seven hundred'	51
czternaście 'fourteen'	46
siedemnaście 'seventeen'	42
jedenaście 'eleven'	42
trzynaście 'thirteen'	40
szesnaście 'sixteen'	39
osiemnaście 'eighteen'	38
dziewiętnaście 'nineteen'	16

Table 9. Polish cardinals ranked with respect to their total frequency
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

In order to confirm the hypothesis that the tripartite division illustrated in Table 3 has been shaped by frequency factors, we would need to show that the three classes of Polish numerals differ considerably in terms of how often the elements that belong to each of them are used. However, as shown in Figure 1, there are no clear-cut frequency thresholds between A-numerals, Q-numerals and N-numerals.

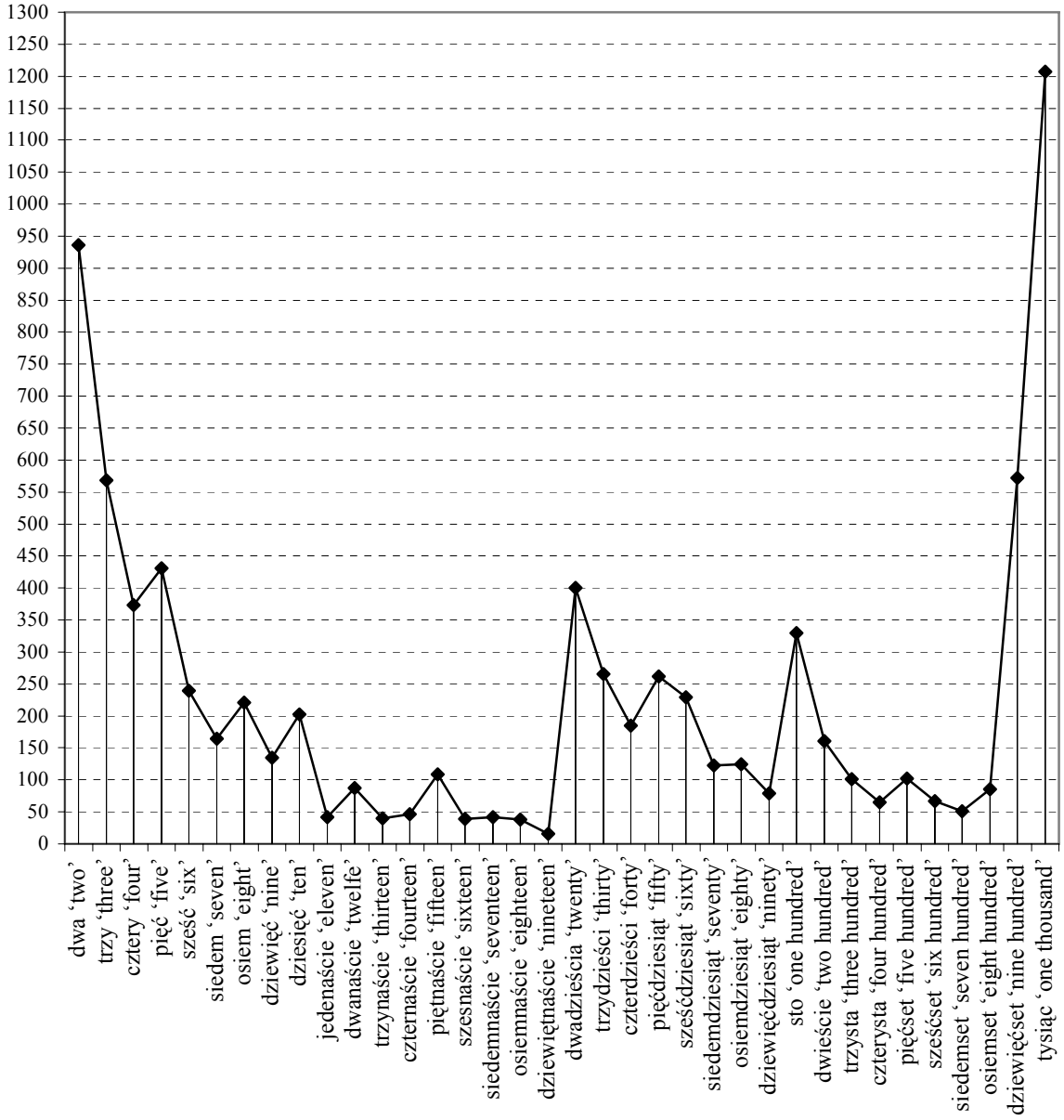

Figure 1. Total frequencies of Polish cardinals
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

The data in Figure 1 support Hammarström's 2004 observation that the numerals for base numbers and their multiples tend to have relatively high frequencies: thus, the numeral for 10 is used more often than the numeral for 9, and the numeral for 100 is significantly more frequent than the numeral for 90. Therefore, if Heine's 1997 model was on the right track, we could expect the words *dziesięć* 'ten' and *sto* 'one hundred' to be more adjectival than *dziewięć* 'nine' and *dziewięćdziesiąt* 'ninety'. However, as shown in Section 2 of the present paper, these frequency differences do not influence the syntax of the numerals in question in any way.

At first glance, the fact that the numerals in the range 11 – 19 are less frequent than the numerals in the range 2 – 9 seems to pattern with Heine's 1997 model (the lower the numeral, the higher its frequency). It should be noted, though, that numerals such as *jedenaście* 'eleven' are expected to have a relatively low frequency because, due to arithmetic factors, their external syntax in complex numeral expressions is very restricted (cf. Gruszczyński and Saloni 1978). As opposed to numerals such as *dwadzieścia* 'twenty', the numerals in the range 11 – 19 cannot be combined with the numerals referring to 1 – 9, which means that there are far fewer complex expressions in which they are used – see (15) vs. (16).

(15) a. dwadzieścia jeden
twenty one
'twenty-one'
(15) b. dwadzieścia dwa
twenty two
'twenty-two'
(15) c. dwadzieścia trzy
twenty three
'twenty-three'
(15) d. dwadzieścia cztery
twenty four
'twenty-four'
(15) e. dwadzieścia pięć
twenty five
'twenty-five'

(16) *jedenaście jeden
eleven one

The expressions in (15a-e) consist of two separate numerals. Every appearance of an expression such as *dwadzieścia dwa* 'twenty-two' has an impact on the total frequencies of both the word *dwadzieścia* 'twenty' and the word *dwa* 'two'. Therefore, contrary to Heine's 1997 model, numerals in the range 20 – 90 are predicted to be more frequent than the numerals in the range 11 – 19. Figure 1 shows that this prediction is borne out.

Apart from total frequency (F), Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 also present data on relative frequency (U) and dispersion (D). The index of dispersion shows the distribution of a particular word in different genres. It ranges between 0 (maximal concentration – an element appears in one genre only) and 100 (the same distribution in all genres). Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 use the indexes F, U and D to divide the lexicon into four strata:

(17) a. grammatical vocabulary (very high frequency and D>80)
(17) b. basic vocabulary (high frequency and 80>D>50)
(17) c. specific vocabulary (F>10, D<50)
(17) d. rare vocabulary (low frequency)

If we apply this classification to numerals, we will see that one numeral lexeme (*dwa* 'two') could be viewed as part of grammatical vocabulary. This patterns with Heine's 1997 assumption that the lowest numerals are adjectival because they have been grammaticalized. However, most other numerals belong to what Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 call basic vocabulary: this class subsumes A-numerals (e.g. *cztery* 'four', D=61,23), Q-numerals (e.g. *osiem* 'eight', D=61,3) and N-numerals (*tysiąc* 'one thousand', D=64,5) – see Table 10. Therefore, there seems to be no reason to view the A/Q/N distinction as related in any way to dispersion.

Numeral	Index D
dwa 'two'	80,5
dziesięć 'ten'	72,16
trzy 'three'	70,81
piętnaście 'fifteen'	66,94
pięć 'five'	66,92
pięćset 'five hundred'	65,18
dwadzieścia 'twenty'	65,15
tysiąc 'thiusand'	64,5
dwanaście 'twelve'	62,68
trzynaście 'thirteen'	62,09
osiem 'eight'	61,3
cztery 'four'	61,23

Table 10. Polish cardinals ranked with respect to the index of dispersion
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

Interestingly, five Q-numerals belong to Kurcz, Lewicki, Sambor, Szafran and Woronczak's 1990 specific vocabulary (Table 11), which suggests that they are more likely to remain nominal. Nevertheless, the dispersion factor has no influence on the syntax of the numerals in question: for instance, there are no syntactic differences between the numeral *piętnaście* 'fifteen' (D=66,94) and the numeral *osiemnaście* 'eighteen' (D=41,88).

Numeral	Index D	Index F
dziewięćdziesiąt 'ninety'	48,96	79
czterysta 'forty'	45,66	65
szesnaście 'sixteen'	44,65	39
osiemnaście 'eighteen'	41,88	38
osiemset 'eight hundred'	39,24	51

Table 11. Q-numerals with the lowest values of the index of dispersion
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

We conclude that the syntactic status of a particular numeral is correlated neither with its total frequency (F), nor with its dispersion in various genres of Polish. Thus, no frequency data seem to support Heine's 1997 model.

Index U (which could be viewed as the index of commonness – cf. Sambor 1972) is calculated by multiplying F by D, in order to relativize the frequency data. When this index is used, the results seem to correspond to Heine's 1997 model to a greater extent than the total frequency data: for instance, the difference between *pięć* 'five' (U=228,43) and *cztery* 'four' (U=228,38) becomes very slight – see Figure 2 (dashed line – total frequencies; solid line – relational frequencies).

However, the relative frequency data are not significantly different from the total frequency data in the case of higher numerals. For instance, the relative frequency of the N-numeral *tysiąc* 'one thousand' is much higher than the relative frequency of the Q-numeral *sto* 'one hundred' – see Table 12. From the point of view of Heine's 1997 model, these data are as problematic as the total frequency data show in Table 9.

Numeral	Index U
sto 'one hundred'	200,63
tysiąc 'one thousand'	778,52

Table 12. Relative frequencies of the numerals for 100 and 1000
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

Finally, it should be noted that Heine's 1997 analysis is also contradicted by the fact that the frequencies of the more conservative (adjectival) forms of the masculine numerals in the range 2 – 4, i.e. *dwaj* 'two', *trzej* 'three' and *czterej* 'four' (see Section 2 of the present paper) are very low – see Table 13.

Numeral	Spoken language	Child language	Popular science	Short press reports	Political commentary journalism	Literary prose	Drama
dwaj 'two'	0	2	2	9	0	4	1
trzej 'three'	0	2	0	4	0	3	0
czterej 'four'	0	3	0	2	0	0	0

Table 13. Total frequencies of the masculine forms *dwaj* 'two', *trzej* 'three' and *czterej* 'four'
(data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990 and Zgółkowa 1983)

These A-type forms are virtually disappearing from Polish. Interestingly, this process is especially salient in colloquial varieties: for instance, forms such as *dwaj* 'two' do not appear in the conversational data. They are being replaced with the Q-type forms *dwóch* 'two', *trzech* 'three', *czterech* 'four', which are obviously less adjectival. This shows that there is no simple correlation between adjectival syntactic properties and high frequency, as proposed by Heine 1997.

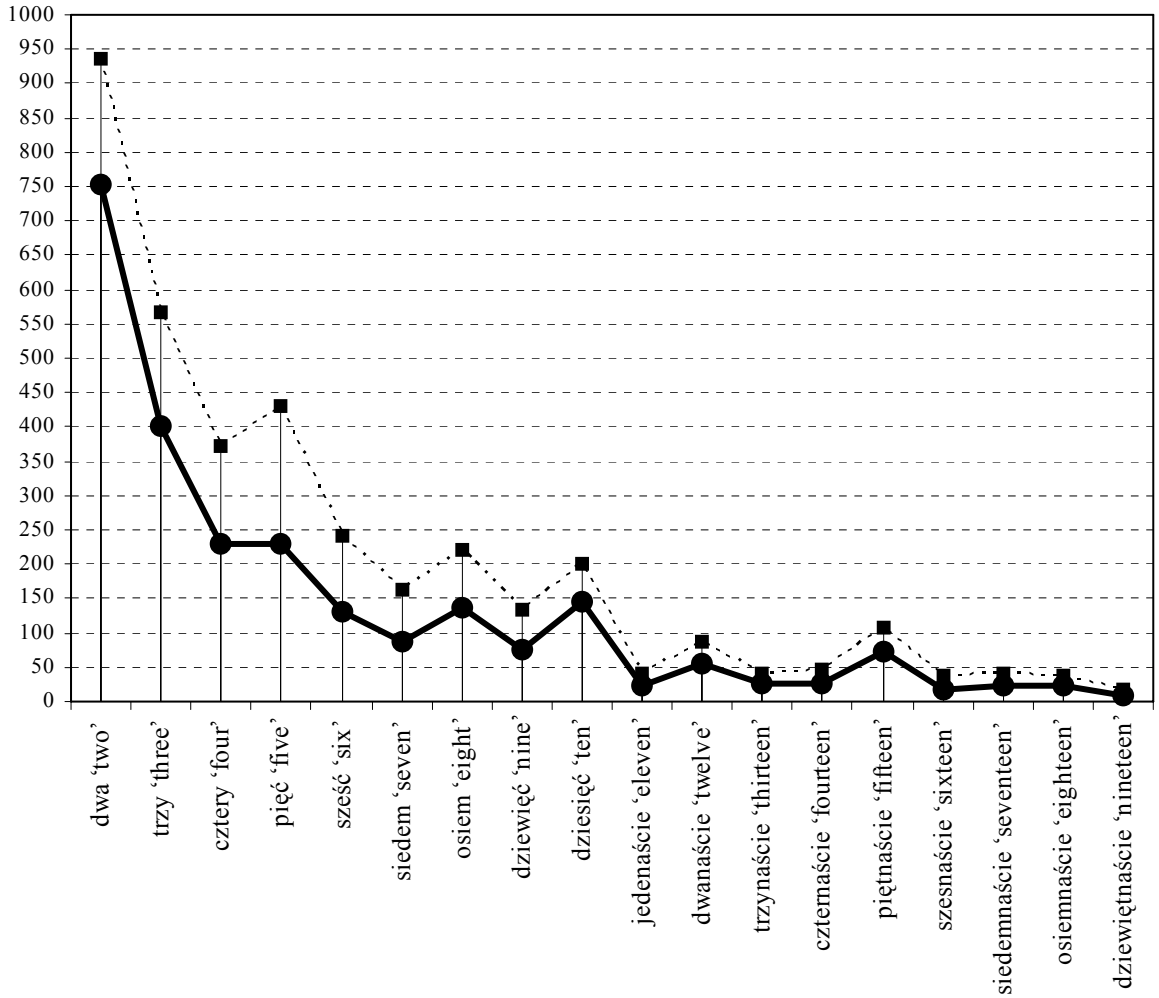

Figure 2. Total and relative frequencies of the numerals *dwa* 'two' – *dziewiętnaście* 'nineteen' (data from Kurcz, Lewicki, Sambor, Szafran and Woronczak 1990)

4 Conclusion

We conclude that Heine's 1997 proposal does not account for the tripartite division of Polish numerals. The data we have analyzed indicate that there is no synchronic correlation between the A/Q/N distinction and frequency. Neither do we find it plausible to assume that, at some stage of the diachronic development of Polish numerals, the frequency of use of the numeral *cztery* 'four' was so much higher than the frequency of the numeral *pięć* 'five' that it caused the adjectival/nominal split. Similarly, the syntactic difference between N-numerals such as *tysiąc* 'one thousand' and Q-numerals such as *dziewięćset* 'nine hundred' does not seem to be related to how often these cardinals are/were used. In other words, we find no reason to assume that the frequency properties of particular numerals shape their syntactic behavior.

References

Corbett, Greville G. 1978. Universals in the Syntax of Cardinal Numerals. *Lingua* 46: 355-368.

Cowan, Nelson. 2001. The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity. Behavioral and Brain Sciences 24: 87-114.

Franks, Steven. 1995. *Parameters of Slavic Morphosyntax*. New York, Oxford: Oxford University Press.

Gruszczyński, Włodzimierz and Saloni, Zygmunt. 1978. Składnia grup liczebnikowych we współczesnym języku polskim. In *Studia gramatyczne II*, ed. Zuzanna Topolińska and Roman Laskowski, 17-42. Wrocław: Ossolineum.
Hammarström, Harald. 2004. Properties of Lower Numerals and Their Explanation: A Reply to Paweł Rutkowski. *Journal of Universal Language* 5.2: 1-20.
Heine, Bernd. 1997. *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
Hurford, James R. 2001. Languages Treat 1 – 4 Specially. *Mind and Language* 16: 69-75.
Juillard, Alphonse G. and Chang-Rodríguez, Eugenio. 1964. Frequency dictionary of Spanish words. The Hague: Mouton.
Kurcz, Ida, Lewicki, Andrzej, Sambor, Jadwiga, Szafran, Krzysztof and Woronczak, Jerzy. 1990. *Słownik frekwencyjny polszczyzny współczesnej*. Kraków: Polska Akademia Nauk, Instytut Języka Polskiego.
Mithun, Marianne. 1999. *The Languages of Native North America*. Cambridge: Cambridge University Press.
Przepiórkowski, Adam. 1996. Case Assignment in Polish: Towards an HPSG Analysis. In *Studies in HPSG* (*Edinburgh Working Papers in Cognitive Science 12*), ed. Claire Grover and Enric Vallduví, 191-228. Edinburgh: University of Edinburgh.
Przepiórkowski, Adam. 2004. O wartości przypadka podmiotów liczebnikowych. *Biuletyn Polskiego Towarzystwa Językoznawczego/Bulletin de la Société Polonaise de Linguistique* 60: 133-143.
Rosch, Eleanor. 1975. Cognitive Reference Points. *Cognitive Psychology* 7: 532-547.
Rutkowski, Paweł. 2000. Składnia polskich grup liczebnikowych: Próba opisu formalnego. *Poradnik Językowy* 8: 10-28.
Rutkowski, Paweł. 2001. Numeral Phrases in Polish and Estonian. In *Proceedings of the 18th Scandinavian Conference of Linguistics*, Vol. 2, ed. Arthur Holmer, Jan-Olof Svantesson and Åke Viberg, 181-190. Lund: Lund University Press.
Rutkowski, Paweł. 2002. Numerals as Grammaticalised Nouns: A Generative Approach. *Interlingüística* 13.III: 317-328.
Rutkowski, Paweł. 2003. On the Universal Neuropsychological Basis of the Syntax of Numerals. *Journal of Universal Language* 4.2: 147-182.
Rutkowski, Paweł. 2006. Grammaticalisation in the Nominal Domain: The Case of Polish Cardinals. Paper presented at the 4th Workshop in General Linguistics, University of Wisconsin, Madison, February 17, 2006.
Sambor, Jadwiga. 1972. *Słowa i liczby. Zagadnienia językoznawstwa statystycznego*. Wrocław: Ossolineum.
Sigurd, Bengt. 1988. Round Numbers. *Language in Society* 17: 243-252.
Zgółkowa, Halina. 1983. *Słownictwo współczesnej polszczyzny mówionej. Lista frekwencyjna i rangowa*. Poznań: UAM.
Zgółkowa, Halina and Bułczyńska, Katarzyna. 1987. *Słownictwo dzieci w wieku przedszkolnym*. Poznań: UAM.

Katedra Językoznawstwa Ogólnego i Bałtystyki
Wydział Polonistyki, Uniwersytet Warszawski
ul. Krakowskie Przedmieście 26/28
00-927 Warsaw, Poland
jlinde@uw.edu.pl
p.rutkowski@uw.edu.pl