

Adriano CERRI

*University of Pisa*

## THE FREQUENCY OF THE USE OF BALTIC NUMERALS: COGNITIVE, LINGUISTIC, AND CULTURAL FACTORS

**Abstract.** This work consists of a corpus-based study on the frequency of the use of numerals in Lithuanian and Latvian. The collected data display the following features: a) the high frequency of the lowest numerals; b) a generally decreasing trend from the first to the last element of each series (units, tens, etc.); c) peaks of frequency often corresponding to important structural positions in the system. In line with the data available for other languages, these features seem to confirm that the universal cognitive abilities and limitations of the speakers play a major role in structuring the frequency of the use of numerals. At the same time, cultural factors are also involved in the process of creating a cognitive hierarchy among number concepts. I suggest distinguishing between innate (cognitive) and non-innate (linguistic and cultural) saliency. The latter is responsible for the identification of reference numbers, thus giving greater credence to the partially relativistic hypothesis that the linguistic system influences the mental mapping of the speakers in the cognitive domain of numbers.

**Keywords:** Baltic; Lithuanian; Latvian; corpus-based study; numerals; frequency; cognitive saliency.

### 1. Introduction

#### 1.1. State of the art

Numerals are often studied from different perspectives: historical, etymological, typological, psychological, etc. Nevertheless, it is rare to come across quantitative studies displaying *how much* and *how often* they are used in a given language. As is well-known, frequency is a relevant parameter in research concerning several topics, such as grammaticalisation (cf. Haspelmath 1999; Dahl 2001; Bybee 2003), lexical diffusion and morphophonological change (cf. Zipf 1929; Mańczak 1969; 2010; Hooper 1976; Berkenfield 2001; Bybee 2002), language acquisition and processing (cf. Ellis 2002; Gülzow, Gagarina 2007; Edwards, Beckman 2008; Divjak,

Gries 2012a; 2012b), etc.<sup>1</sup> However, quantitative approaches are seldom applied to the study of numerals. Some examples are Mańczak (1985, 348) on the frequency of Italian numerals and Hurford (1987, 90–92) on English. Sigurd (1988) compares Swedish and English numerals' frequency counts focusing on “round” numbers (e.g. 10, 20, 25, 30, 50, 100, etc.); he shows that the *roundness* of a number depends on the base of the system and it can be calculated through a formula. So far, the most comprehensive study in this field is Dehaene, Mehler (1992) where cross-linguistic regularities in the frequency of numerals are detected and explained with particular reference to psychological factors. In their study, seven unrelated languages are taken into account: American English, Catalan, Dutch, French, Japanese, Kannada, Spanish.

Conversely, no similar research exists for the Baltic languages, i.e., Lithuanian and Latvian. So far, Baltic numerals have been mostly described and studied according to the historical-comparative method.<sup>2</sup> Therefore, it can be worthwhile taking up this issue adopting a quantitative methodology.

### 1.2. Basic assumptions and objectives

In this work a linguistic expression shall be considered a *numeral* if and only if (cf. McGregor 2014, 50):

- (1) a. it is a lexical item (simple or complex);
- b. its meaning specifies a numerical quantity;
- c. it can occur in a syntagmatic relation with an entity-specifying lexeme.

So, non-lexical numerical expressions such as *two times four* or *the square root of 64* must be left out (see 1.a); quantifiers like *many*, *some*, *a few* are also to be excluded (see 1.b). The same holds for fractions, which do not express an absolute but a relative numerical value.<sup>3</sup> Point (2.c) leads us to leave out

---

<sup>1</sup> For a wide overview of the main tendencies and the results of the quantitative approach in linguistics, see Köhler, Altmann, Piotrowski (2005); more specifically on the frequency of use, see Bybee, Hopper (2001), Baayen (2001), Bybee (2007).

<sup>2</sup> To name only the main works: Endzelīns (1923, 356–372; 1957, 141–146), Mažiulis (1957; 1965, 604–636), Sokols et al. (1959, 477–500), Stang (1966, 276–285), Zinkevičius (1981, 56–70; 1996, 133–138), Smoczyński (1986), Comrie (1992), Forssman (2001, 141–157), Pokrotniece (2002, 243–308).

<sup>3</sup> Lithuanian grammars traditionally include the class of *trupmeniniai* (fractions) in the category of numerals, cf. Ambrazas 2005; Paulauskienė 2007.

the class of the so-called “collective” Latvian numerals. This class contains five items formed by the cardinal stem followed by the suffix *-atā* (an old locative): *divatā*, *trijatā*, *četratā*, *piecatā*, *sešatā*. They cannot quantify an entity-specifying lexeme (a noun, for instance). Rather, they modify a verb, e.g. *dzīvot trijatā* literally ‘to live in three’, *mēs strādājam četratā* literally ‘we work in four’. Therefore they should be considered adverbs.

Before proceeding, it is also useful to recall some general notions on numerals. Numeral systems share regularities and common features among the languages of the world (cf. Stampe 1976; Corbett 1978; Greenberg 1978; Seiler 1990; Comrie 1997; 2005). A basic feature is that they are not merely sequences of number words, but they belong to a structure. In other words, every single element, being part of a system, has a specific status and plays a role: “numeral systems are not only linearly, but also hierarchically ordered” (Gvozdanović 1992, 8).

Let us consider the most common type of numeral system, i.e., the decimal one. Here we have: *a*) simple units: 1, 2, 3...; *b*) the main base: 10; *c*) powers of the base: 100, 1.000...; *d*) upper units: 20, 30, 40...; 200, 300, 400...; 2.000, 3.000, 4.000... In addition, there can be secondary bases like 60 in French: *soixante-dix* ‘seventy’ (literally ‘sixty-ten’), or 20 in Irish: *trí fichid* ‘sixty’ (literally ‘three twenty’), sub-bases, traces of older bases, “magic” or culturally-prominent numbers, and so on. Given this situation, an analysis of the frequency of the use of numerals will help to address the following questions:

1. Are Lithuanian and Latvian numerals used with a random frequency?
2. If not, what kind of pattern of distribution emerges?
3. Do Lithuanian and Latvian data show noticeable differences?
4. Is there any connection between the data on the frequency of use and the structure of the numeral systems?
5. How to account for the frequency pattern?

As I am going to motivate below, it is reasonable to assume that analysing the usage of numerals can help us to understand the mental mapping of speakers in the specific cognitive domain of numbers. In order to do this, it will suffice to consider a very simple parameter, the absolute frequency (F), i.e., the total number of occurrences of a type in a corpus.<sup>4</sup> More sophisti-

---

<sup>4</sup> The inflected forms must be considered variables (tokens) of one single numeral (type).

cated parameters, such as relative frequency, mean frequency, rank, dispersion, etc. are not necessary in this case because I am not going to compare the frequency of an item with its counterpart in the other language, nor to examine the variations of frequency in different portions of the corpora.

## 2. Instruments and Methods

### 2.1. Corpora and data extraction

My data are taken from the two largest monolingual corpora available for the target languages, namely the Corpus of Contemporary Lithuanian (*Dabartinės lietuvių kalbos tekstynas*, henceforth – DLKT)<sup>5</sup> and the Balanced Corpus of Contemporary Latvian Texts (*Līdzsvarotais mūsdienų latviešu valodas tekstu korpuss*, henceforth – LVK2018).<sup>6</sup> Tables 1–2 report information on these instruments, such as their dimensions and composition, their subfields, and so on. Unfortunately, neither of the corpora take spoken language into account.<sup>7</sup> Of course, it would have been possible to use other, more specific databases, such as the Corpus of Academic Lithuanian (CorALit: <<http://coralit.lt>>), or one of the Latvian corpora available on the platform <<http://www.korpuss.lv>>. However, DLKT and LVK2018 were preferred because of the bigger amount and the heterogeneity of their data. These features permit to get a more representative picture of the current, neutral usage of the two languages, thus reducing – as far as possible – the impact of contingent factors (e.g. diaphasic, diatopic or diastratic variation) on the frequency of a linguistic item.

Table 1. **Composition of DLKT** (cf. <<http://tekstynas.vdu.lt/tekstynas/menu?page=statistics>>)

Textual typology	No. of words	Percentage
Prose (literary)	18.461.597	11,6 %
Prose (non-literary)	21.024.249	14,2 %
Administrative	13.625.715	10,0 %

<sup>5</sup> <[tekstynas.vdu.lt/tekstynas](http://tekstynas.vdu.lt/tekstynas)>.

<sup>6</sup> <[korpuss.lv/id/LVK2018](http://korpuss.lv/id/LVK2018)>.

<sup>7</sup> By the time I carried out the present research, no corpora of the spoken language were available. Luckily, at least for Lithuanian, this gap has recently been filled by the publication of the *Sakytinės lietuvių kalbos tekstynas* (Corpus of spoken Lithuanian language: <[sakytinistekstynas.vdu.lt/index.php](http://sakytinistekstynas.vdu.lt/index.php)>).

Textual typology	No. of words	Percentage
Journalism	87.251.905	63,8 %
Spoken language	557.822	0,3 %
Total no. of words	140.921.288	100,0 %

Table 2. **Composition of LVK2018**  
(cf. <http://www.korpuss.lv/id/LVK2018>)

Textual typology	No. of words	Percentage
Periodicals	5.850.047	60 %
Literature	2.025.875	20 %
Scientific literature	986.929	10 %
Normative acts	733.497	7 %
Acts of Parliament	216.663	2 %
Total no. of words	9.813.014	100 %

A noticeable difference between these two tools is that the Latvian corpus is (automatically) morphologically annotated, while the Lithuanian is not. This fact determines different data extraction procedures for the two languages. For Latvian I simply carried out queries by headword, while for Lithuanian I had to search for all the possible forms of each numeral<sup>8</sup> and then I added up the occurrences. With complex numerals, such as *divdešimt penki* ‘25’, I used the function *išplėstinė paieška* “expanded search” and I selected *frazė* “sentence”.

## 2.2. The samples

Below, the samples for the two languages are shown. (Note that the existing studies on other languages do not usually take the following classes into account: 21–29; hundreds; cardinals for *pluralia tantum*; collectives; ordinals 10<sup>th</sup>–90<sup>th</sup>; ordinals 10<sup>th</sup>–100<sup>th</sup>–1.000<sup>th</sup>.)

Lithuanian sample:

- I. Simple cardinals: I.a) simple units 1–9; I.b) tens 10–90; I.c) 11–19; I.d) 21–29; I.e) hundreds 100–900; I.f) powers of the base 10<sup>3</sup>, 10<sup>6</sup>, 10<sup>9</sup>
- II. Cardinals for *pluralia tantum* (*dauginiai*)

<sup>8</sup> Of course, not only the standard forms, but also the colloquial ones, such as those with shortened endings for dative (*-iem*, *-iom* instead of *-iems*, *-iom(i)s*) and for locative (*-iuos*, *-ios* instead of *-iuose*, *-iose*).

III. Collectives (*kuopiniai*)

IV. Ordinals: IV.a) 1<sup>st</sup>–9<sup>th</sup>; IV.b) 10<sup>th</sup>–90<sup>th</sup>; IV.c) 11<sup>th</sup>–19<sup>th</sup>; IV.d) 100<sup>th</sup>, 1.000<sup>th</sup>, 1.000.000<sup>th</sup>, 1.000.000.000<sup>th</sup>

Latvian sample:

I. Simple cardinals: I.a) simple units 1–9; I.b) tens 10–90; I.c) 11–19; I.d) 21–29; I.e) hundreds 100–900; I.f) powers of the base 10<sup>3</sup>, 10<sup>6</sup>, 10<sup>9</sup>

II. Cardinals for *pluralia tantum*

III. Ordinals: III.a) 1<sup>st</sup>–9<sup>th</sup>; III.b) 10<sup>th</sup>–90<sup>th</sup>; III.c) 11<sup>th</sup>–19<sup>th</sup>; III.d) 100<sup>th</sup>, 1.000<sup>th</sup>, 1.000.000<sup>th</sup>, 1.000.000.000<sup>th</sup>

### 3. Data

The collected data are presented in the charts. It was found to be more convenient to present the results sub-dividing the material into different series: 1–9 (Charts 1–2), the tens (Charts 3–4), the powers of the base (Charts 5–6), 200–900 (Charts 7–8), 11–19 (Charts 9–10), 21–29 (Charts 11–12), cardinals for *pluralia tantum* (Charts 13–14), ordinals 1<sup>st</sup>–9<sup>th</sup> (Charts 15–16), ordinals 10<sup>th</sup>–90<sup>th</sup> (Charts 17–18), ordinals 11<sup>th</sup>–19<sup>th</sup> (Charts 19–20), ordinals for the powers of 10 (Charts 21–22), Lithuanian collectives (Chart 23). Solid lines in the charts show the general tendency; they help to make the general orientation of the values more evident. A quick look at the charts will suffice to reveal recurring patterns of distribution.

### 4. Discussion

#### 4.1. General features

Some general features of the collected data are immediately recognisable. One of the main ones is a decreasing trend from the first to the last element of each cycle. Another peculiarity is the higher frequency of some numerals compared to others. High frequencies are denoted by “peaks”. It is noteworthy that peaks often coincide with numerals holding an important structural position such as the main base (10), upper units (20, 30) and the powers of the base (100, 1000). Finally, let me point out that in both languages numerals 1, 2, 3 and, to a lesser extent, 4 are much more frequent than all the others.

To sum up, the collected data are characterised by three main features:

- (2) a. the high frequency of the very first numerals 1-2-3-(4);
- b. a decreasing trend from the first to the last element of each cycle;
- c. peaks of frequency often corresponding with important structural positions in the system.

Chart 1. Lithuanian numerals 1-9

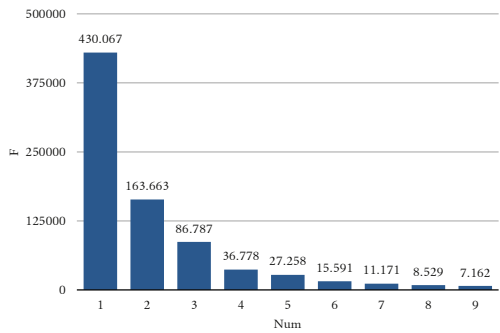


Chart 2. Latvian numerals 1-9

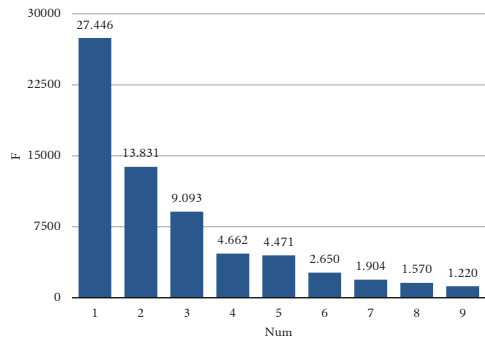


Chart 3. Lithuanian tens 10-90

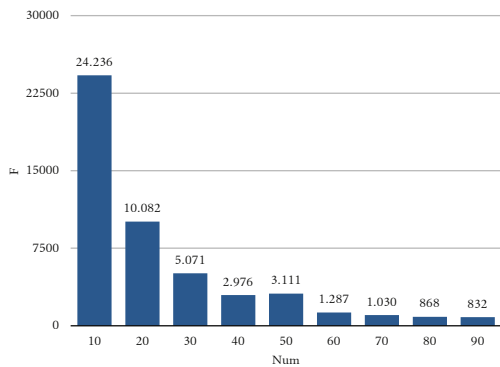


Chart 4. Latvian tens 10-90

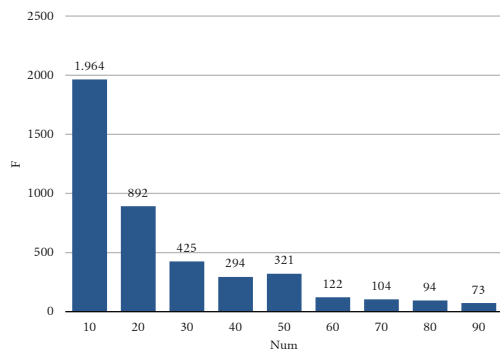


Chart 5. Lithuanian powers of 10

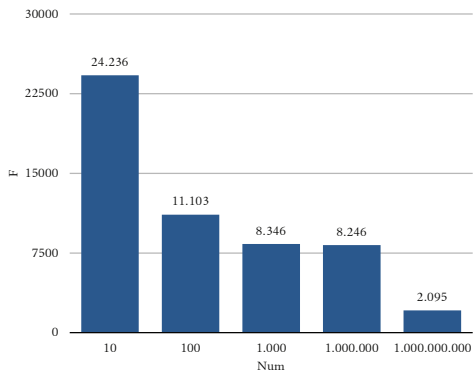


Chart 6. Latvian powers of 10

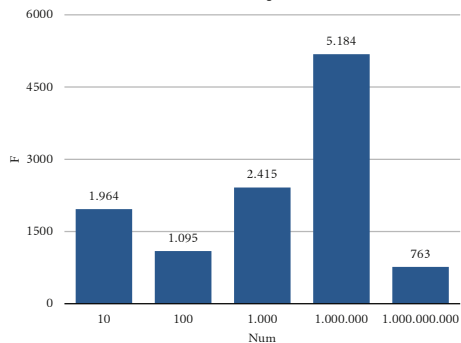


Chart 7. Lithuanian hundreds 200–900

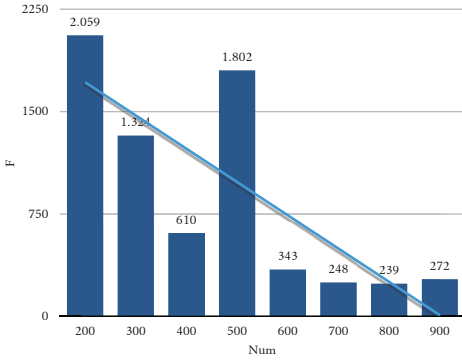


Chart 8. Latvian hundreds 200–900

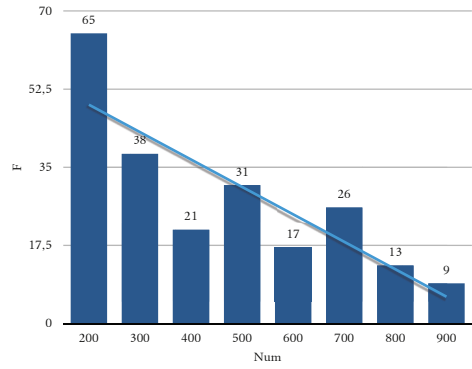


Chart 9. Lithuanian numerals 11–19

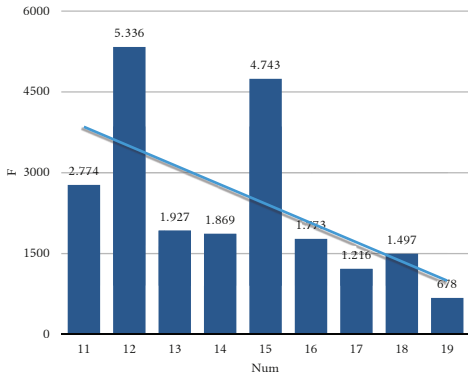


Chart 10. Latvian numerals 11–19

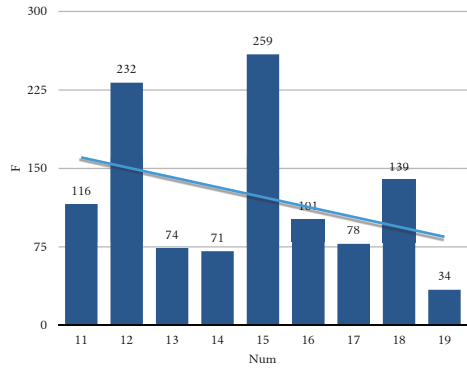


Chart 11. Lithuanian numerals 21–29

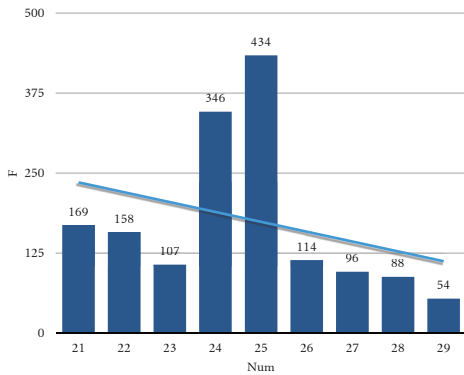


Chart 12. Latvian numerals 21–29

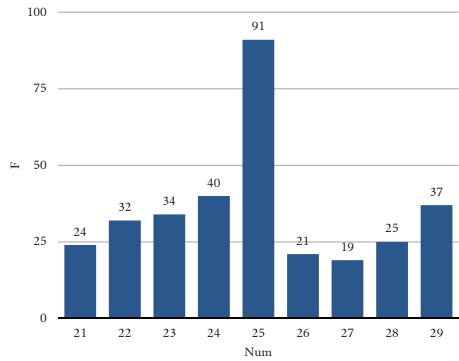




Chart 13. Lithuanian cardinals for pluralia tantum (dauginiai)

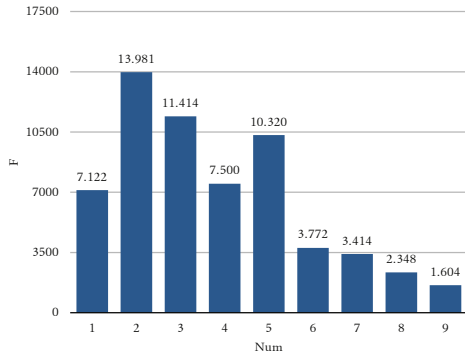


Chart 14. Latvian cardinals for pluralia tantum

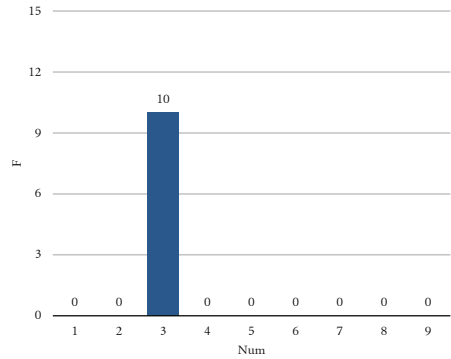


Chart 15. Lithuanian ordinals 1<sup>st</sup>-9<sup>th</sup>

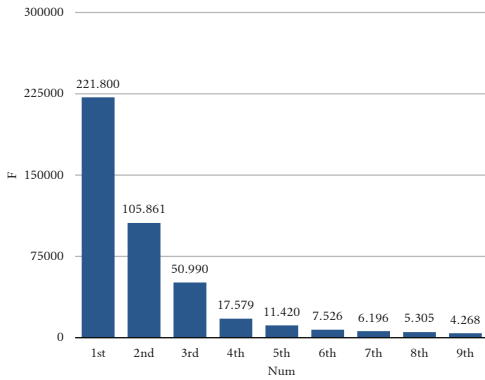


Chart 16. Latvian ordinals 1<sup>st</sup>-9<sup>th</sup>

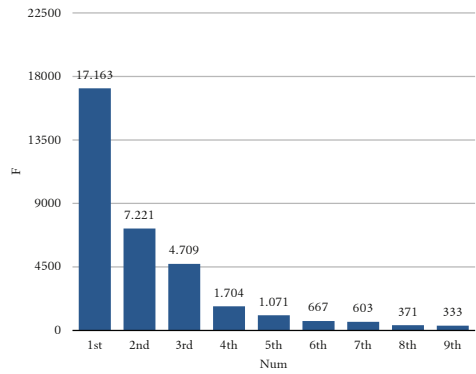


Chart 17. Lithuanian ordinals 10<sup>th</sup>-90<sup>th</sup>

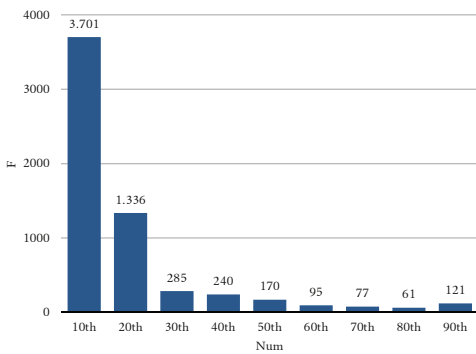


Chart 18. Latvian ordinals 10<sup>th</sup>-90<sup>th</sup>

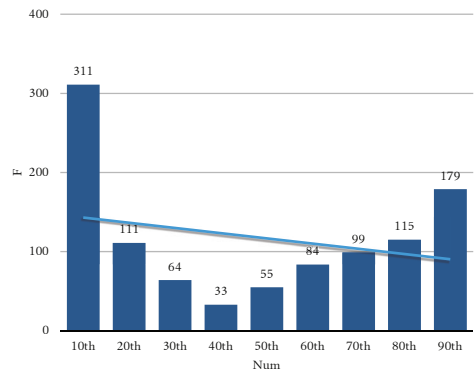


Chart 19. Lithuanian ordinals 11<sup>th</sup>-19<sup>th</sup>

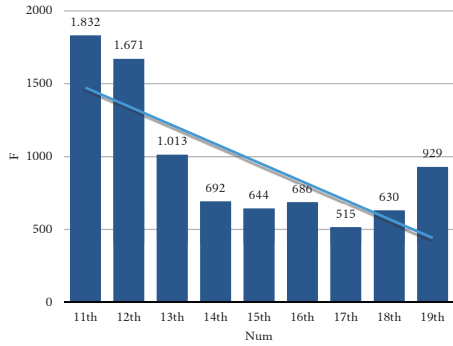


Chart 20. Latvian ordinals 11<sup>th</sup>-19<sup>th</sup>

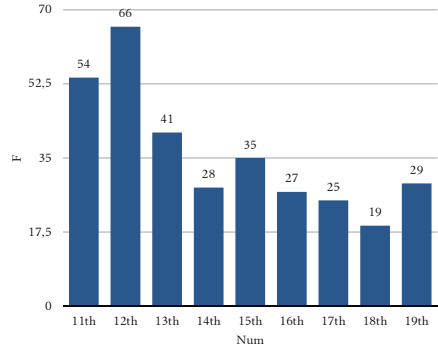


Chart 21. Lithuanian ordinals for the powers of 10

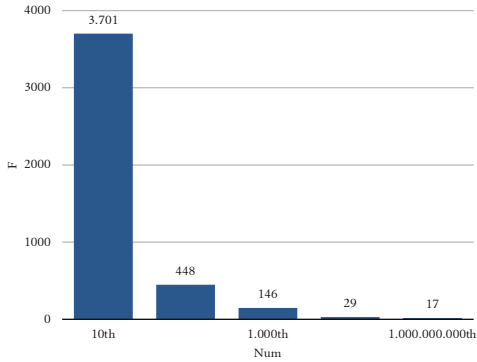


Chart 22. Latvian ordinals for the powers of 10

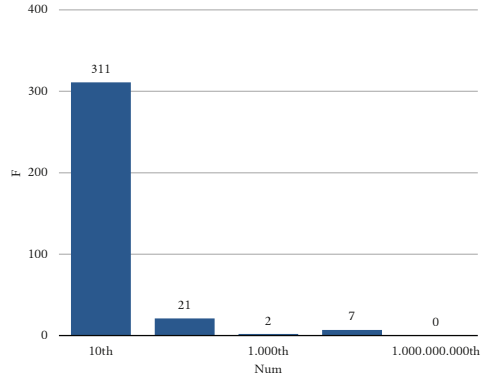
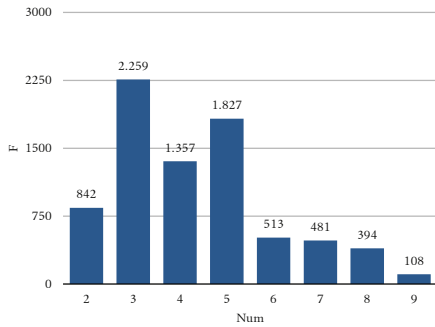


Chart 23. Lithuanian collectives (kuopiniai)



These results provide an answer to our starting questions:

1. Are Lithuanian and Latvian numerals used with a random frequency? *No, they are not.*
2. If not, what kind of pattern of distribution emerges? *A pattern of the type described in (2).*
3. Do Lithuanian and Latvian data show noticeable differences? *There are only small differences, while the features listed in (2) hold for both languages.*
4. Is there any correlation between the frequency data and the structure of the numeral systems? *Yes. This is shown by the correspondence between peaks of frequency and round numbers.*

The regularity of (2) is so specific and recurring that it can hardly be due to a coincidence. If we also consider that similar results are also found in other languages (see below), it becomes self-evident that this is not a random result. So, I will now focus on the major question: how can we account for the frequency pattern described in (2)?

#### 4.2. The decreasing trend

A macroscopic feature of the collected data is the great prevalence of Lith. *vienas*, Latv. *viens*. Actually, the word indicating ‘one’ in Lithuanian and Latvian – as well as in many other languages – is also used as a pronoun and an adjective. Since the Lithuanian corpus is non-annotated, only a part of those occurrences refers to the numeral. Hence, the occurrences of Lith. *vienas* are overestimated. Nevertheless, this does not prevent me from drawing the conclusion that *vienas* is by far the most frequently used numeral in Lithuanian. Furthermore, this is also confirmed by the two main frequency dictionaries of Modern Lithuanian (Grumadienė, Žilinskienė 1997; Utkā 2009); see Table 3:

Table 3. **Different sources show that Lith. *vienas* ‘1’ is the most frequent numeral**

	Grumadienė, Žilinskienė (1997)	Utkā (2009)	DLKT
Num	F	F	F
<i>vienas</i> ‘1’	2.641	1.257	430.067
<i>du</i> ‘2’	1.557	1.177	163.663

<i>trys</i> '3'	738	544	86.787
<i>keturi</i> '4'	276	249	36.778
<i>penki</i> '5'	187	156	27.258
<i>šeši</i> '6'	122	92	15.591
etc.			

As mentioned above, a very high frequency characterises not only 'one', but also 'two' and 'three'. How can we account for this data? In my opinion, a satisfactory explanation comes from cognitive psychology. Since the late 1940s, many studies have shown that humans (as well as some animal species) are able to recognise immediately small cardinalities – up to six items according to Kaufman, Lord, Reese, Volkman (1949), up to four according to Cowan (2000), or even two/three according to Fischer (1992). This “number sense” (cf. Dehaene 1997) is called *subitizing*. It seems to be innate and it differs from *counting*, which is a more complex and non-innate ability.<sup>9</sup> Subitizing reveals that the human mind creates a hierarchy among number concepts.<sup>10</sup> Thus, lower cardinalities are more salient and recognisable than higher ones: one is more salient than two, which is in turn more salient than three, and so on. Since what is more salient can also be expected to be more frequent in language, subitizing possibly accounts for the high frequency of the lowest numerals.

Now let us focus on the series 1–9 (see Charts 1–2). We have observed a clear trend: the frequency decreases as the numerical value increases. This is very consistent with the results of Dehaene, Mehler (1992); my data confirm this cross-linguistic regularity. In this case subitizing is not a sufficient explanation, being restricted only to the first three/four cardinalities. Dehaene, Mehler (1992, 17–18) suggest considering Fechner’s law, according to which:

---

<sup>9</sup> Subitizing is fast (< 100 ms/item) and accurate; counting is accurate, though, much slower (> 200 ms/item) than subitizing, and it can be observed with a larger set of objects (more than four items), cf. Krajcsi, Szabó, Mórocz (2013, 227).

<sup>10</sup> On subitizing see also Mandler, Shebo (1982), Dehaene, Cohen (1994), Trick, Pylyshyn (1994), Simon, Vaishnavi (1996), Peterson, Simon (2000), Cerri (2013).

small numbers receive an expanded and more accurate mental representation relative to large numbers. [...] Assuming that the frequency of production of numerals is directly related to the importance of the associated mental number representation or number concept, a decrease in frequency with magnitude is predicted.

This gives us an insight into our minds; here the concept of number is structured according to a pattern: the lower the number, the more accurate its mental representation. Note that the essence of this explanation is the same as for subitizing: again, what is more salient (and/or accurately represented) can also be expected to be more frequent in the linguistic production. Therefore, I agree in considering Fechner's law a good explanatory model.

On the one hand, laboratory experiments show that lower numbers correlate with a more accurate mental representation (i.e. with a higher cognitive saliency). On the other hand, quantitative analyses (including the present one) show that lower numbers correlate with a higher frequency. Hence the syllogistical conclusion that frequency (when not due to contingent factors) correlates with cognitive saliency. Therefore it is reasonable to interpret the former as a clue for the latter.

Not only for numerals 1–9 does the decreasing trend hold true; indeed, we find it in upper series as well. The series of the tens is very similar to that of the simple units (compare Charts 3–4); something similar is seen for the hundreds (Charts 7–8). The powers of 10 behave in a similar way (see Charts 5–6); however, it should be pointed out that here Latv. *tūkstoš* / *tūkstoš* 'one thousand' and *miljons* 'one million' do actually represent exceptions. The same trend is also observed in the distribution of ordinals (Charts 15–22; the only exceptions are Latv. 50<sup>th</sup>–90<sup>th</sup>).

In both languages 11–19 show some interesting features: though there is a decreasing trend, there are also some noticeable exceptions in relation to 12 and 15 (see Charts 9–10). These exceptions become explicable if we consider some important cultural factors such as the partition of time. Lith. *penkiolika* '15' (F = 4.743) often occurs in the phrases *penkiolika metų* '15 years' (1.481 times, 31,2 %) and *penkiolika minučių* '15 minutes' (533 times, 11,2 %). In fact, the latter is a culturally accepted portion of time, which is more salient (and more frequently used) than 14 or 16 minutes. If we carry out a similar search for the phrase *keturiolika minučių* '14 minutes' we obtain a very different result: only in 11 cases (0,58 %) the word *keturiolika* co-occurs

with *minučių*. The use of ‘15’ in time expressions with ‘years’ is even more frequent (in relative terms) in Latvian. *Piecpadsmīt* ( $F = 259$ ) co-occurs with *gads* 114 times (44 %).

A similar explanation can be argued for the high frequency of 12. Firstly, the cultural importance of this number is confirmed by the existence of a specific lexeme in many languages: It. *dozzina*, Sp. *docena*, Fr. *douzaine*, Engl. *dozen*, Germ. *Dutzend*, Lith. *tuzinas*, Latv. *ducis*, Rus. *djúžina*, Hung. *tucat*, Turk. *düzine*, etc. The number 12, though it is rarely used as a base in the world’s languages, often holds a special status in the numeral systems, at least among the languages of Europe. In the Germanic world, the presence of the *Grosshundert* ‘120’ and the *Grosstaused* ‘1.200’, i.e., the “long” (duodecimal) hundred and thousand, is well attested. The presence of such non-decimal values has led some scholars to hypothesise the historical interference of a base 12. Nevertheless, this idea cannot be accepted. As Sommer (1951, 65) points out, if 12 were a real base, then its multiple would be 144 ( $12 \times 12$ ), not 120. The “long” Germanic hundred/thousand should be regarded as a case of *semantical reinterpretation* of the base due to material, culturally-determined circumstances. In particular, the reinterpretation of the Indo-European decimal ‘hundred’ as 120 seems to have originated from fish-trading in the North Atlantic and the Baltic Sea: fish and other goods were traded in stocks of 12/120 units, where the remaining 2 or 20 represented a margin of discount (cf. Seiler 1990, 194). In our culture 12 owes its importance first of all to its role in the measuring of time. A year is divided into 12 months, a day into two cycles of 12 hours, 12 is a factor of 60 (the minutes of one hour), etc. Moreover, this number is popular in traditional Lithuanian folklore (cf. Rūķe-Draviņa 1979, 54). This leads us to two considerations: firstly, beyond these exceptions, the above-mentioned trend is confirmed. Secondly, to account for the high frequency of 12 or 15 we do not have to resort to psychological arguments (such as subitizing or Fechner’s law), but to cultural ones.

It is interesting to look at the series 21–29. As far as I know, these numerals have never been considered in previous studies. While in the Baltic languages, as well as in many other languages, 11–19 are formed by single number words (synchronically speaking), 21–29 are complex numerals, similarly to 31–39, 41–49, etc. As a consequence, the series 21–29 can be more useful than 11–19 as a model for the following series. My data show that in

this series there is also a decreasing trend, and least in Lithuanian, though it is much less evident than in the lower series; moreover, there are interruptions in relation to 24 and 25 (see Chart 11).<sup>11</sup> Also in this case, cultural motivations account for the frequency of 24 (the hours of the day) and 25 (a quarter of 100). As an example, the query *dividėšimt keturias* in the DLKT produces 124 strings. In 102 cases (82,2 %) the numeral is followed by *valandas* ‘hours’.

Beyond this, the frequency of the use follows the same decreasing trend recognisable elsewhere (see the tendency line in Chart 11). If we accept the idea that Fechner’s law is at work here (our mental representation is more accurate for smaller rather than for bigger numbers), we have to think that its effect tends to become more negligible with higher cardinalities (the more we go on, the less the difference is significant). The distribution of the data seems to be of a logarithmic type. As is shown in Table 4, the range of occurrences (from the highest to the lowest F) in each series diminishes more and more – e.g. in Lithuanian, from hundreds of thousands to thousands in the series 1–9, from thousands to hundreds in the series 11–19, and from hundreds to tens in the series 21–29. Something similar is seen in Latvian.

Table 4. **Range of F in the series 1–9, 11–19 and 21–29**

Series	Lith.		Latv.	
	Highest F	Lowest F	Highest F	Lowest F
1–9	430.067	7.162	27.446	1.220
11–19	5.336	678	259	34
21–29	434	54	91	19

These data could represent the linguistic reflection of a cognitive architecture. The effects of Fechner’s law are evidently mirrored by the different F of, say, 3 as compared with that of 8; conversely, the difference between F(23) and F(28) is not so relevant because it corresponds with a weakly differentiated accuracy in the mental representation of the number concepts 23 and 28.

---

<sup>11</sup> Latvian data are not considered here because the counting of these numerals produced too few occurrences (mean frequency = 35.8); so, the distribution might not be significant (see Chart 12).

Now I would like to make a few observations on the cardinals for *pluralia tantum* (see Charts 13–14). As far as the Latvian ones are concerned, it is self-evident that these forms have almost completely disappeared from the everyday language. Though they are still registered in the grammars, in real contexts of use *vienēji*, *dīvēji*, *treji*, etc. are replaced by the simple cardinals *viens*, *divi*, *trīs*, etc. This is the main observable difference between Lithuanian and Latvian. As a matter of fact, the class of Lithuanian *dauginiai* is actively used by the speakers, though here as well there is a tendency – especially in the spoken language and in the colloquial register – to replace the *dauginiai* with simple cardinals, e.g. *jau gal trys metai jau, kai mes susirašinėjam* (‘perhaps we have been writing to each other for three years’); *Pastarieji penki metai...* (‘The last five years...’). According to the normative rules of standard Lithuanian, *treji* and *penkeri* should be used instead of *trys* and *penki* respectively.

It should also be pointed out that the most frequently attested form is not *vieneri* ‘1’, but *dveji* ‘2’. This can be explained from a historical point of view. While the forms for *pluralia tantum* from *dveji* ‘2’ to *devyneri* ‘9’ are old, *vieneri* is a quite recent formation due to the analogical effect of 2–9 on 1 (cf. Murinienė 1997). So, in many cases the oldest norm (i.e., the use of the simple cardinal *vieni*) is still used. In this case, the decreasing trend starts from the second element of the series.

### 4.3. Peaks of frequency

The data show that 10 is much more frequent than 9 or 11; the same holds for 20 as compared with 19 or 21 and for many other cases. These breaks in the decreasing trend are meaningful. It is very likely that for a speaker of a decimal language 10 is much more than just “the number after 9”. The collected data demonstrate this fact by displaying a big leap from 9 to 10 (see Table 5). A similar leap is found between the end of each series and the beginning of the following one:

Table 5. Frequency “leaps” between the end of a series and the beginning of a new one

	F(9)	F(10)	F(19)	F(20)	F(90)	F(100)
Lithuanian	7.162	24.236	678	10.082	832	11.103
Latvian	1.220	1.964	34	892	73	1.095



We have already noticed that peaks of frequency often correspond to those numerals which hold an important structural position. This means that the data on frequency somehow match with the internal structure of the numeral system. In particular, peaks of frequency correspond with round numerals, especially the base and its multiples. Therefore

- (3) a correspondence is shown between the frequency of the use of a numeral, its structural role, and its cognitive saliency.

We now need to clarify the mutual relationship between such different aspects. In order to do this, two main questions must be addressed: why are there peaks of frequency? And why those peaks (and not others)? The following reflections will broaden the discussion to a cross-linguistic level: hopefully, our data on the Baltic languages allow for conclusions which can also apply to other languages.

I believe that within the distribution of frequency, peaks reveal a psychological necessity. Speakers need “milestones” along the endless path of numbers. In fact, according to mathematics and arithmetic, we can describe numbers as a uniform sequence of elements which are formed by the addition of one unit:  $N^* = \{1, 1+1, 1+1+1, 1+1+1+1, \dots\}$ . However, such a formal representation is of little use in everyday life: our cognitive limitations do not allow us to deal with such a shapeless, uniform sequence. Thus, our mind needs to create a pattern, to give a shape and a hierarchy to those elements. Hence the need for “reference numbers” (cf. Dehaene, Mehler 1992, 18). The underlying pattern is similar to that of the “cognitive reference points” illustrated by Eleanor Rosch (cf. especially Rosch 1975 where the domain of numbers is studied).

Another reason for the presence of peaks of frequency is that some numerals are used for approximations. Out of the many occurrences of 20, a part of them refers to a quantity of approximately, rather than exactly 20 elements. Conversely, this is not possible for 19 or 21. So, I can say *a twenty-year-old boy* meaning ‘a young man’ in a broad sense; instead, if I say *a nineteen-year-old boy*, I am referring to his precise age. Similarly, round numbers are also used in indefinite contexts of the type *I have told you a thousand times...* or *I have a million things to do today* (cf. Luján Martínez 1995, 216). Approximate and indefinite uses are only possible with reference numbers.

We can ask ourselves what actually determines cognitive reference numbers. Are they universal or are they determined by the language (or, more in general, by the culture) of the speakers? If we hypothesise that they are cognitive universals like subitizing, we need to conclude that our mind assigns a different status to different cardinalities. According to this view, the structure of a numeral system would be a *consequence* of the cognitive saliency of some number concepts (e.g. 10, 20, 100, etc.) which are *intrinsically* more salient than others. Since there seems to be nothing intrinsically peculiar in the cardinality of, say, 10 as compared to that of 9, I find it more reasonable to hypothesise that cognitive reference points are determined by the language (and by the culture). In this view, the decimal structure of a numeral system is not a consequence, but *the cause* of the cognitive saliency of some number concepts. This interpretation leads to a mildly-relativistic hypothesis:

- (4) cognitive reference numbers are (at least in part) determined by the language of the speakers.

Peaks of frequency correspond with relevant points because the structure of the system itself provides a special status to some numerals. This fulfils our need for reference points and enables us to handle such a complex tool as the sequence of numbers. Starting from universal cognitive limitations, speakers need cognitive reference points, and these are chosen – at least in part – on the basis of the linguistic structure they use. So, in this area, language would influence the cognitive structure of the speakers. A different mapping of the cognitive domain of numbers may correspond with different linguistic structures.

One may object that what determines the mental mapping is not the language, but more broadly the culture in its various aspects. Lithuanian and Latvian cultures are largely decimal: decimal scales are used for measuring weights, dimensions, money, and so on.<sup>12</sup> So, the presence of a decimal base

---

<sup>12</sup> However, it should be pointed out that instances of non-decimal counting were evident in the Baltic world, especially in the traditional rural culture, until some decades ago. Attestations of this are such lexical relics as Lith. *kapa*, Latv. *kapa* ‘60, threescore’ (from a Slavic \**kopa* originally meaning ‘heap’, later used as a commercial term for a batch of sixty items, cf. Comrie 1992, 781); Lith. *colis* ‘inch’ (2,54 cm); Lith. *pūdas*, Latv. *puds* ‘unit of measurement for weight corresponding to 16,38 kg’ (< ORu. *пудъ*), etc. (cf. LKŽ<sup>e</sup>, LLVV<sup>e</sup>, *s.v.*). Though they all seem to be due to linguistic/cultural contact, it would be interesting to study the state of vitality of these non-decimal elements.

in the numeral system may reflect the cultural customs adopted by the society in such activities like counting and measuring. The interplay between linguistic and cultural aspects in this specific area has not been clarified yet and surely deserves more attention.<sup>13</sup> For the moment, it can be said that language *and* culture determine the cognitive mapping of the speakers in the domain of numbers.

## 5. Generalisations

### 5.1. Innate vs. non-innate saliency

In view of the above considerations, I suggest distinguishing between *innate* and *non-innate* saliency in the conceptual domain of numbers. The former refers to the properties of the human mind to recognise and represent certain cardinalities; this has to do with the mental representation of number concepts (see subitizing and Fechner's law). The latter is strictly connected with the linguistic structure of the system. Our mind does not assign more saliency to, say, 19 or 20 items. There is no intrinsic property of the cardinality itself that makes one of these two number concepts more salient than the other; what actually gives a peculiar saliency to 20 is the structure of the numeral system (in our case the decimal structure of the Lithuanian and Latvian systems). Whereas innate saliency is natural and universal,<sup>14</sup> non-innate saliency is cultural, i.e., acquired through learning and experience, see Figure 1.

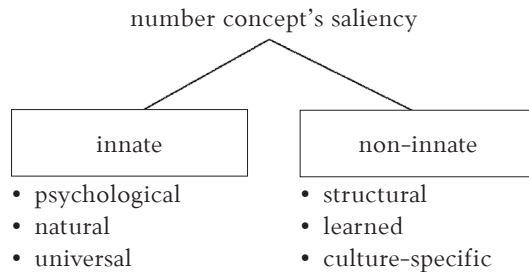


Figure 1. **Innate and non-innate saliency**

---

<sup>13</sup> Cf. Everett (2013): for an outline on this topic (pp. 61–71); and for more evidence of the influence of numerical language on the conceptualisation of quantities (pp. 140–169).

<sup>14</sup> The universality of innate saliency is supported by the cross-linguistic and cross-cultural regularities reported by Dehaene, Mehler (1992).

A brief remark on non-innate saliency. If it is culturally-determined, we can expect it to exhibit a high degree of variability. In fact, the languages of the world show a variety of numeral systems such as quaternary, quinary, decimal, duodecimal, vigesimal, sexagesimal, etc.<sup>15</sup> Nevertheless, the range of choice is not completely arbitrary, as all the different systems seem to be historically based on the same cognitive foundation, i.e. the *body-part model* (cf. Seiler 1990, 193; Heine 1997, 19–24). For instance, four corresponds to the fingers of one hand with the exception of the counting thumb, five and ten correspond to the fingers of one or two hands, twelve corresponds to the phalanxes of one hand (again, with the exception of the counting thumb), and so on.<sup>16</sup> Hence, within this culturally-determined area of variability, a major role is played by the body-part model (see also below).

### 5.2. Cognitive, linguistic and cultural factors

In order to account for our data it was necessary to recall several arguments involving three main classes of factors: *cognitive*, *linguistic* and *cultural*. In this last part of the discussion, I will try to provide an overall picture of these factors and to highlight their mutual relationship.

Cognitive factors, part of what I have called the innate saliency, include those that influence the mental representation of number concept and the limitations that drive the speakers to make use of reference numbers.

Linguistic and cultural factors form the non-innate saliency, which includes the linguistic structure of the numeral system: lexical units, bases, good-formation rules which permit the creation of higher numerals, etc. Cultural aspects refer to the way a society counts objects, measures time, distances, money, and so on. Everyday experience shows that, for instance, non-decimal numbers can have a peculiar status in a decimal-language-speaking society (that is the case of the measuring of time in 60 seconds, 60 minutes, 12+12 hours, etc.). One can also imagine a different kind of cultural influence on the usage of numerals, i.e., folk traditions, taboos, superstitions, etc. Such aspects can restrain, inhibit or, conversely, foster the frequency of some numerals. As a matter of fact, this aspect is not mirrored in our data. For instance, though 9 is a special, highly symbolic number in Baltic and Slavic

---

<sup>15</sup> The most comprehensive collection of numeral systems is provided by the website *Numeral Systems of the World's Languages*, <[mpi-lingweb.shh.mpg.de/numeral/](http://mpi-lingweb.shh.mpg.de/numeral/)> (= Chan 2013).

<sup>16</sup> See Ifrah (1994) and Justus (2004) on this.

folklore (cf. Polívka 1927; Ozols 1993, 84), it is less frequent than 7 or 8. This fact is not surprising since traditional literature is poorly represented in our corpora (cf. Table 1–2).

Cognitive, linguistic and cultural factors should not be intended as closed classes. They interact at both the synchronic and the diachronic level. Innate cognitive skills allow humans to *a)* distinguish a few small cardinalities and *b)* represent number concepts with a decreasing accuracy. These are the basics of number reasoning, but they are still not sufficient for a full mastery of the number domain. Such a mastery needs to rely on reference numbers. Therefore, these are identified on the basis of the non-innate saliency (see Figure 2). In other words, speakers of different speech communities choose those numbers which are significant in accordance with their numeral system (linguistic factors) and their social conventions (cultural factors).

It is more difficult to clarify the historical and functional interplay between language and culture: do we count on base  $n$  because we speak an  $n$ -base language, or vice versa? I would like to suggest a third possibility. In this context, we have to abandon the idea of a perfect correspondence between language and cultural practices. I have already referred to decimal languages adopting instances of non-decimal counting and measuring.<sup>17</sup> The converse is also attested in the case of populations speaking non-decimal languages, nevertheless using decimal counting – at least as a consequence of the contact with Western culture. So, if these two aspects are partly independent, they probably do not relate in terms of cause-and-effect. Rather, we have to consider the third possibility of their common origin.

Historical, archaeological and anthropological evidence demonstrates that human abilities in the field of number have developed over millennia from concrete practices of digital counting (cf. Menninger 1969; Schmandt-Besserat 1992; Chrisomalis 2009). Mankind learned to count – that is to conceptualise and to name numbers – with the assistance of concrete supports. The “discovery” of number as an abstract property equating, say, three children and three pebbles, was reached only after a long phase in which the counted entities were paired with entities taken from a reference-cluster (e.g., stones, sticks, fingers, etc.). The human body has always been the most practical reference-cluster. It is easy to find plentiful evidence of etymologi-

---

<sup>17</sup> Several examples can be found in Justus (1999).

cal connections between number names and body parts, especially “hand”, “finger” or similar, in many languages of the world, including the Indo-European family.<sup>18</sup> This indicates that in ancient times number names were often coined in accordance with the corresponding body part. So the relationship between counting practices and numeral systems should be understood as a history of mutual influence grounded on the same cognitive foundation, the body-part model.

By contrast, other cultural practices such as time-measuring, numerical superstitions, taboos, etc. cannot always be assigned to the influence of the body-part model. This makes the interplay between linguistic and cultural factors worthy of further exploration.

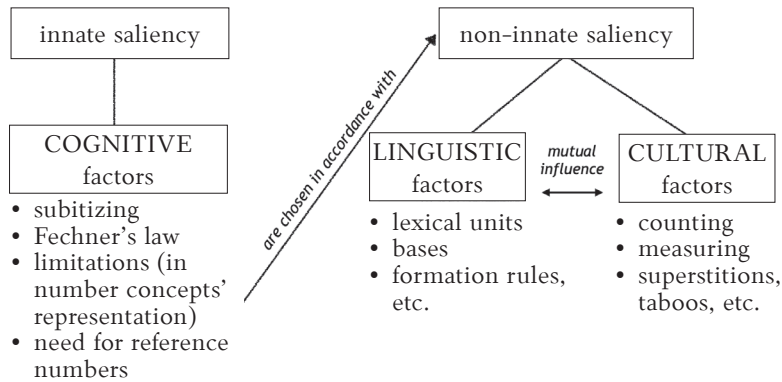


Figure 2. **Cognitive, linguistic and cultural factors influencing number concept saliency**

## 6. Conclusions

To sum up, I would like to recall the main results of the present study. The data extracted from the two major corpora of written Lithuanian and Latvian show that Baltic numerals are not used with random frequency. For both

<sup>18</sup> The literature on this point is very extensive. Transparent etymological connections between body parts and numerals are particularly frequent among the languages of Africa (cf. Zaslavsky 1979), as well as in the Austro-Asiatic (cf. Rischel 1997) and in the Austronesian (cf. Lean 1986–1988) language families. The Indo-European languages are less transparent; however, analogous connections have been suggested for them, too. For example, Szemerényi (1960, 69) reconstructs for *ide*. \**de-k̑mt* 10 the original meaning ‘two hands’ (\**kont-* ‘hand’, cf. Gmc. \**χanduz*).

languages it is possible to identify a recurring pattern of distribution characterised by three main features: *a*) the high frequency of the first three numerals; *b*) a generally decreasing trend from the first to the last element of each cycle; *c*) peaks of frequency often corresponding with important structural positions in the system (e.g. the base, its multiples, “round” numbers, etc.).

As far as the first point is concerned, I highlighted the role of *subitizing*; as an innate property of the human mind, it provides a special cognitive saliency for the first three/four cardinalities. Since what is more salient can also be expected to be more frequent in language, subitizing seems to be a good explanation for the high frequency of the lowest numerals. Note that this is an internal explanation that gives priority to the cognitive property of the speakers. Other scholars prefer adopting an external (pragmatic, referential) explanation; according to this view, the lowest numerals are used more frequently because in everyday situations the need to refer to a low number *n* is likely to arise more often than a need to refer to a bigger number (cf. Hurford 1987, 90ff).

The second feature of the recurring pattern of distribution can be expressed as follows: within each series, the lower the numeral, the higher its frequency of use, and vice versa. This result is very consistent with what has already been observed for other languages. Interestingly, my data show that such a pattern can be interpreted as a basic model which is also recognisable in the upper series (e.g. 11–19, 10–90, 100–900, etc.) as well as in the different classes (ordinals, collectives, numerals for *pluralia tantum*). In order to account for the regular decreasing trend I accept the idea suggested by Dehaene, Mehler (1992) that this feature is also due to some cognitive properties of the human mind, in this case Fechner’s law. The accuracy of the mental representation of number concepts decreases as the magnitude increases. So, the smaller the number, the more accurate its mental representation. Consequently, the frequency of use decreases with larger numerals (within each series).

Exceptions to this trend are rare: that is the case, in both Lithuanian and Latvian, of 12, 15, 24, 25. These cases are only explicable by considering cultural, rather than cognitive, factors. Culturally-determined practices (though adopted all over the world) like the partition of time (12 months, 24 hours, etc.) or the need for convenient divisors (15 minutes as a quarter of an hour, 25 as a quarter of 100, etc.) influence the use of numerals.

Lithuanian and Latvian differ only with respect to the class of numerals for *pluralia tantum*. According to my data, this class has fallen into disuse in Latvian (see Chart 14). The role of these numerals has almost been completely absorbed by simple cardinals. On the other hand, numerals of this class are still in use in Lithuanian – at least in the written language, while they are diminishing in the spoken language.

Another noticeable result is that peaks of frequency often correspond with those numerals which hold an important structural position in the decimal system. A correspondence is shown between the structural role of the numeral, its cognitive saliency, and its frequency of use. In general, the presence of peaks of frequency can be accounted for by the theory of cognitive reference points (cf. Rosch 1975). Speakers need “milestones” which help them to create a mental mapping for the cognitive domain of number. Peaks correspond with number concepts which are especially salient with respect to others. We have questioned what determines the choice of reference numbers in Baltic as well as in other languages. I have argued that cognitive reference points are chosen on the basis of what I have called the non-innate saliency, i.e. by linguistic and cultural factors. If one could consider the linguistic effects alone, it would be possible to suggest the relativistic claim that language influences the mental structure of the speakers in this area of cognition. So, different mapping of the cognitive domain of numbers may correspond with different numeral systems. The former, in turn, should be reflexed in the frequency of use.<sup>19</sup> But the present state of the art only permits the formulation of a mildly-relativistic version of this hypothesis: cognitive reference num-

---

<sup>19</sup> It would be worthwhile developing this idea, especially adopting a cross-linguistic perspective. It would surely be of interest to analyse non-decimal languages in order to test the present results. For instance, does a quinary language display a pattern of decreasing frequency? If yes, is it repeated in the other series (e.g. 6–9, 11–14)? Do the frequency peaks match with the quinary structure? Or one could consider the case of a vigesimal language: does the decreasing trend proceed from one to 19, or do our cognitive limitations require a lower reference point? Unfortunately, strong limitations for this kind of studies are the small number of purely non-decimal languages, and the scant availability of adequate tools (online corpora, reliable databases, softwares, etc.) for such languages. Moreover, as Comrie (2005, 531) states, “Nondecimal numeral systems are even more endangered than the languages in which they occur” because decimal systems tend to be adopted as a consequence of the cultural contact with the Western world.



bers are determined by the speakers' language and culture. As a last remark, I would like to recall that both cultural practices (counting) and linguistic structures (numeral systems) are historically grounded on a common cognitive foundation, the body-part model.

## **BALTŲ KALBŲ SKAITVARDŽIŲ VARTOJIMO DAŽNUMAS: KOGNITYVINIAI, LINGVISTINIAI IR KULTŪRINIAI FAKTORIAI**

### *Santrauka*

Straipsnyje pristatomas tekstynų medžiaga paremtas lietuvių ir latvių kalbų skaitvardžių vartojimo dažnumo tyrimas. Surinkta medžiaga rodo šiuos požymius: a) didelis mažiausius skaičius įvardijančių skaitvardžių dažnumas; b) bendra dažnumo mažėjimo tendencija kiekvienos serijos (vienetų, dešimčių ir t. t.) elementų eilėje nuo pirmo iki paskutinio; c) dažnumo viršūnės dažnai atitinka svarbias struktūrinės pozicijas sistemoje. Kartu su kitų kalbų duomenimis šie požymiai patvirtina, kad kalbos vartotojų universalūs kognityviniai gebėjimai ir jų ribotumas atlieka svarbų vaidmenį skaitvardžių dažnumo distribucijos struktūroje. Tuo pačiu metu kultūriniai faktoriai yra taip pat įtraukti į skaičių konceptų kognityvinės hierarchijos kūrimo procesą. Straipsnyje siūloma skirti įgimtą (kognityvinį) ir neįgimtą (lingvistinį ir kultūrinį) iškilumą. Pastarasis yra atsakingas už referencinių skaičių identifikaciją – taip suteikiama daugiau patikimumo iš dalies reliatyvistinei hipotezei, kad kalbos sistema daro įtaką kalbėtojų mentalinio žemėlapio formavimui skaičių kognityvinėje srityje.

## **REFERENCES**

Ambrazas, Vytautas (ed.) 2005, *Dabartinės lietuvių kalbos gramatika* (4 pataisytas leidimas), Vilnius: Mokslo ir enciklopedijų leidybos institutas.

Baayen, Harald R. 2001, *Word frequency distributions*, Dordrecht, Boston, London: Kluwer Academic Publishers.

Berkenfield, Catie 2001, The role of frequency in the realization of English 'that', in Joan L. Bybee, Paul J. Hopper (eds.), *Frequency and emergence of linguistic structure*, Amsterdam: John Benjamins, 281–308.

Bybee, Joan L. 2002, Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change, *Language Variation and Change* 14, 261–290.

Bybee, Joan L. 2003, Mechanisms of change in grammaticization: the role of frequency, in Brian D. Joseph, Richard D. Janda (eds.), *The handbook of historical linguistics*, Oxford: Blackwell, 602–623.

Bybee, Joan L. 2007, *Frequency of use and the organization of language*, Oxford: Oxford University Press.

Bybee, Joan L., Paul J. Hopper (eds.) 2001, *Frequency and emergence of linguistic structure*, Amsterdam: John Benjamins.

Cerri, Adriano 2013, Riflessi linguistici del subitizing, *Studi e saggi linguistici* 51(1), 53–81.

Chan, Eugene S. L. 2013, *Numeral systems of the world's languages*, Hong Kong, <https://mpi-lingweb.shh.mpg.de/numeral/> (2019 11 05).

Chrisomalis, Stephen 2009, The cognitive and cultural foundations of numbers, in Eleanor Robson, Jacqueline Stedall (eds.), *The Oxford handbook of the history of mathematics*, Oxford: Oxford University Press, 495–517.

Comrie, Bernard 1992, Balto-Slavonic, in Jadranka Gvozdanović (ed.), *Indo-European numerals*, Berlin, New York: Mouton de Gruyter, 717–833.

Comrie, Bernard 1997, Some problems in the theory and typology of numeral systems, in Bohumil Palek (ed.), *Proceedings of the LP '96*, Prague: Karolinum, 41–56.

Comrie, Bernard 2005, Numeral bases, in Martin Haspelmath, Matthew S. Dryer, David Gil, Bernard Comrie (eds.), *The world atlas of language structures*, Oxford: Oxford University Press, 530–532.

Corbett, Greville G. 1978, Universals in the syntax of cardinal numerals, *Lingua* 46, 355–368.

Cowan, Nelson 2000, The magical number 4 in short-term memory: a reconsideration of mental storage capacity, *Behavioral and Brain Sciences* 24(1), 87–114.

Dahl, Östen 2001, Inflationary effects in language and elsewhere, in Joan L. Bybee, Paul J. Hopper (eds.), *Frequency and Emergence of Linguistic Structure*, Amsterdam: John Benjamins, 471–480.

Dehaene, Stanislas, Jacques Mehler 1992, Cross-linguistic regularities in the frequency of number words, *Cognition* 43, 1–29.

Dehaene, Stanislas, Laurent Cohen 1994, Dissociable mechanisms of subitizing and counting: neuropsychological evidence from simultanagnosic patients, *Journal of Experimental Psychology: Human Perception and Performance* 20, 958–975.

Dehaene, Stanislas 1997, *The number sense: how the mind creates mathematics*, New York: Oxford University Press.

Divjak, Dagmar S., Stefan T. Gries (eds.) 2012a, *Frequency effects in language learning and processing*, Berlin: Mouton de Gruyter.

Divjak, Dagmar S., Stefan T. Gries (eds.) 2012b, *Frequency effects in language representation*, Berlin: Mouton de Gruyter.

DLKT – *Dabartinės lietuvių kalbos tekstynas* (Corpus of contemporary Lithuanian), <http://donelaitis.vdu.lt>.

Edwards, Jan, Mary E. Beckman 2008, Some cross-linguistic evidence for modulation of implicational universals by language-specific frequency effects, *Language Learning and Development* 4(2), 122–156.

Ellis, Nick C. 2002, Frequency effects in language processing. A review with implications for theories of implicit and explicit language acquisition, *Studies in Second Language Acquisition* 24(2), 143–188.

Endzelīns, Jānis 1923, *Lettische Grammatik*, Heidelberg: Carl Winter.

Endzelīns, Jānis 1957, *Baltų kalbų garsai ir formos*, Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla.

Everett, Caleb 2013, *Linguistic relativity. Evidence across languages and cognitive domains*, Berlin, Boston: de Gruyter Mouton.

Fischer, Jean-Paul 1992, Subitizing: the discontinuity after three, in Jacqueline Bideaud, Claire Meljac, Jean-Paul Fischer (eds.), *Pathways to number: children's developing numerical abilities*, Hillsdale: Lawrence Erlbaum, 191–208.

Forssman, Berthold 2001, *Lettische Grammatik*, Dettelbach: J. H. Röhl.

Greenberg, Joseph H. 1978, Generalizations about numeral systems, in Joseph H. Greenberg, Charles A. Ferguson, Edith A. Moravcsik (eds.), *Universals of human language 3: Word structure*, Stanford: Stanford University Press, 249–295.

Grumadienė, Laima, Vida Žilinskienė 1997, *Dažninis dabartinės rašomosios lietuvių kalbos žodynas (mažėjančio dažnio tvarka)*, Vilnius: Lietuvių kalbos institutas, Matematikos ir informatikos institutas.

Gülzow, Insa, Natalia Gagarina (eds.) 2007, *Frequency effects in language acquisition: Defining the limits of frequency as an explanatory concept*, Berlin, New York: Mouton de Gruyter.

Gvozdanović, Jadranka (ed.) 1992, *Indo-European numerals*, Berlin, New York: Mouton de Gruyter.

Haspelmath, Martin 1999, Why is grammaticalization irreversible?, *Linguistics* 37(6), 1043–1068.

Heine, Bernd (1997), *Cognitive foundations of grammar*, New York, Oxford: Oxford University Press.

Hooper, Joan B. (1976), Word frequency in lexical diffusion and the source of morphophonological change, in William M. Christie (ed.), *Current progress in historical linguistics*, Amsterdam: North Holland, 95–105.

Hurford, James R. 1987, *Language and number: The emergence of a cognitive system*, Oxford: Basil Blackwell.

Ifrah, Georges 1994, *Histoire universelle des chiffres*, Paris: Lafont.

Justus, Carol F. 1999, Pre-decimal structures in counting and metrology, in Jadranka Gvozdanović (ed.), *Numeral types and changes worldwide*, Berlin, New York: Mouton de Gruyter, 55–79.

Justus, Carol F. 2004, On language and the rise of a base for counting, *General Linguistics* 42, 17–43.

Kaufman, E. L., M. W. Lord, T. W. Reese, J. Volkmann 1949, The discrimination of visual number, *American Journal of Psychology* 62(4), 498–525.

Köhler, Reinhard, Gabriel Altmann, Rajmund G. Piotrowski (eds.) 2005, *Quantitative Linguistik / Quantitative Linguistics*, Berlin, Paris: de Gruyter.

Krajcsi, Attila, Eszter C. Szabó, István Á. Mórocz 2013, Subitizing is sensitive to the arrangement of objects, *Experimental Psychology* 60(4), 227–234.

Lean, Glendon A. 1986–1988, *Counting systems of Papua New Guinea* 1–17 (draft edition), Port Moresby: Papua New Guinea University of Technology.

LKŽ<sup>o</sup> – Gertrūda Naktinienė, Jonas Paulauskas, Ritutė Petrokienė, Vytautas Vitkauskas, Jolanta Zabarskaitė (red.), *Lietuvių kalbos žodynas* 1–20 (1941–2002), electronic version, Vilnius: Lietuvių kalbos institutas (www.lkz.lt).

LLVV<sup>o</sup> – *Latviešu literārās valodas vārdnīca* (Dictionary of the literary Latvian language), LU Matemātikas un informātikas institūts, LU Latviešu valodas institūts, www.tezaurs.lv/llvv/.

Luján Martínez, Eugenio R. 1995, Los usos no numerales de las palabras para «uno» y la raíz indoeuropea \*sem, *Cuadernos de Filología Clásica* 5, 215–230.

LVK2018 – *Līdzsvaroats mūsdienu latviešu valodas tekstu korpus* (Balanced Corpus of Contemporary Latvian Texts), korpus.lv/id/LVK2018 (last access: 2019/11/05).

Mańczak, Witold 1969, *Le développement phonétique des langues romanes et la fréquence*, Kraków: Uniwersytet Jagielloński.

Mańczak, Witold 1985, Indo-European numerals and the sexagesimal system, in Jacek Fisiak (ed.), *Papers from the Sixth Conference on Historical Linguistics*, Amsterdam, Poznań: John Benjamins, 347–352.

Mańczak, Witold 2010, Irregular sound change due to frequency in Germanic languages, *Lingua Posnaniensis* 52(1), 55–65.

Mandler, George, Billie J. Shebo 1982, Subitizing: an analysis of its component processes, *Journal of Experimental Psychology: General* 111(1), 1–22.

Mažiulis, Vytautas 1957, Apie lietuvių kalbos skaitvardžius, in Chackelis Lemchenas (ed.), *Kai kurie lietuvių kalbos gramatikos klausimai*, Vilnius: Valstybinė politinės ir mokslinės literatūros leidykla, 40–65.

Mažiulis, Vytautas 1965, Skaitvardis, in Kazys Ulvydas (ed.), *Lietuvių kalbos gramatika 1: Fonetika ir morfologija*, Vilnius: Mintis, 604–636.

McGregor, William B. 2014, Numerals and number words in Shua, *Journal of African Languages and Linguistics* 35(1), 45–90.

Menninger, Karl 1969, *Number words and number symbols: a cultural history of numbers*, Cambridge: MIT Press.

Murienė, Lina 1997, Ar klaida: penki tvartai, vienos durys?, *Gimtoji kalba* 10, 7–8.

Ozols, Arturs 1961 [1993], *Latviešu tautasdziesmu valoda*, Rīga: Latvijas valsts izdevniecība.

Paulauskienė, Aldona 2007, Skaitvardis, in Pranas Kniūkšta (ed.), *Lietuvių kalbos žinynas*, Kaunas: Šviesa, 135–141.

Peterson, Scott A., Tony J. Simon 2000, Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture, *Cognitive Science* 24, 93–122.

Pokrotņiece, Kornēlija (ed.) 2002, *Latviešu literārās valodas morfoloģiskās sistēmas attīstība*, Rīga: Latvijas universitāte, Latviešu valodas institūts.

Polívka, Jiří 1927, Les nombres 9 et  $3 \times 9$  dans les contes des slaves de l'est, *Revue des Études Slaves* 7(3/4), 217–223.

Rischel, Jørgen 1997, Typology and reconstruction of numeral systems: The case of Austroasiatic, in Jacek Fisiak (ed.), *Linguistic reconstruction and typology*, Berlin, New York: Mouton de Gruyter, 273–312.

Rosch, Eleanor 1975, Cognitive reference points, *Cognitive Psychology* 7(4), 532–547.

Rūķe-Draviņa, Velta 1979, On numerals in Baltic and Slavic languages, *Acta Baltico-Slavica* 12, 53–66.

Schmandt-Besserat, Denise 1992, *Before writing 1: From counting to cuneiform*, Austin: University of Texas Press.

Seiler, Hansjakob 1990, A dimensional view on numeral systems, in William Croft, Keith Denning, Suzanne Kemmer (eds.), *Studies in typology and diachrony (for Joseph Greenberg)*, Amsterdam, Philadelphia: John Benjamins, 187–208.

Sigurd, Bengt 1988, Round numbers, *Language in Society* 17(2), 243–252.

Simon, Tony J., Sandeep Vaishnavi 1996, Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages, *Perception and Psychophysics* 58, 915–926.

Smoczyński, Wojciech 1986, Bałto-słowiańskie ordinalia na tle indoeuropejskim, *Acta Baltico-Slavica* 17, 199–231.

Sokols, Evalds et al. (eds.) 1959, *Mūsdienu latviešu literārās valodas gramatika 1: Fonētika un morfoloģija*, Rīga: Latvijas PSR zinātņu akadēmijas izdevniecība.

Sommer, Ferdinand 1951, *Zum Zahlwort*, München: Bayerische Akademie der Wissenschaften.

Stampe David 1976, Cardinal number systems, in Salikoko S. Mufwene, Carol A. Walker, Sanford B. Steever (eds.), *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, Chicago: Chicago Linguistic Society, 594–609.

Stang, Christian 1966, *Vergleichende Grammatik der baltischen Sprachen*, Oslo, Bergen, Tromsø: Universitetsforlaget.

Szemerényi, Oswald J. L. 1960, *Studies in the Indo-European system of numerals*, Heidelberg: Carl Winter.

Trick, Lana M., Zenon W. Pylyshyn 1994, Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision, *Psychological Review* 101(1), 80–102.

Zaslavsky, Claudia 1973 [1979], *Africa counts: number and pattern in African culture*, Brooklyn, New York: Lawrence Hill Books.

Zinkevičius, Zigmas 1981, *Lietuvių kalbos istorinė gramatika 2*, Vilnius: Mokslas.

Zinkevičius, Zigmas 1996, *The history of the Lithuanian language*, Vilnius: Mokslo ir enciklopedijų leidykla.

Zipf, George K. 1929, Relative frequency as a determinant of phonetic change, *Harvard Studies in Classical Philology* 40, 1–95.

*Adriano CERRI*

*Università di Pisa*

*Piazza Evangelista Torricelli 2*

*I-56126 Pisa*

*Italy*

[[adriano.cerri@fileli.unipi.it](mailto:adriano.cerri@fileli.unipi.it)]