

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The version of this text has not yet been defined or was untraceable and may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/74350>

Please be advised that this information was generated on 2013-07-29 and may be subject to change.



---

## On Round Numbers: Pragmatic Aspects of Numerical Expressions\*

C.J.M. Jansen<sup>1</sup> and M.M.W. Pollmann<sup>2</sup>

<sup>1</sup>Nijmegen University, The Netherlands

<sup>2</sup>Utrecht University, The Netherlands

---

### ABSTRACT

This paper describes and explains some regularities in the frequency of numbers in text. An analysis of number frequencies in text corpora in Dutch, English, German, and French confirms the expectation that frequency is highly dependent on two factors: magnitude and roundness. Roundness (defined as number frequency in an approximation context) proves to be related to three arithmetical properties: '10-ness', '2-ness', and '5-ness'. In predicting the frequency of numbers irrespective of their context '2½-ness' should be added to these factors, as is suggested in the work of Sigurd (1988). The role of the four number characteristics found in this study can be explained by the preference of the language user for using base numbers, and for doubling and halving quantities.

### INTRODUCTION

Obviously, there are number frequency differences in language use. Not all numbers occur equally frequently. It is easy to predict that the number 2 is used more often than the number 20, 20 more often than 24, and 24 more often than 244. We collected frequency data for number tokens from the so-called 27-mln-corpus of the Institute for Dutch Lexicology in Leiden. It consists of the editorial text of the leading Dutch newspaper *NRC/Handelsblad*, from January, 1994 up to and including April, 1995. The frequencies for the numbers 2-1000 that we found in this Dutch text corpus are shown in Figure 1. For reasons of transparency, a logarithmic scale is used for the numbers on the *x*-axis.

---

\*Address correspondence to: C.J.M. Jansen, Center for Language Studies, Department of Business Communication, Nijmegen University, Erasmus Building 5.23, PO Box 9103, 6500 HD Nijmegen, The Netherlands. Tel.: +31-24-3612045; Fax: +31-24-3612177; E-mail: c.jansen@let.kun.nl; Website: <http://www.careljansen.nl>

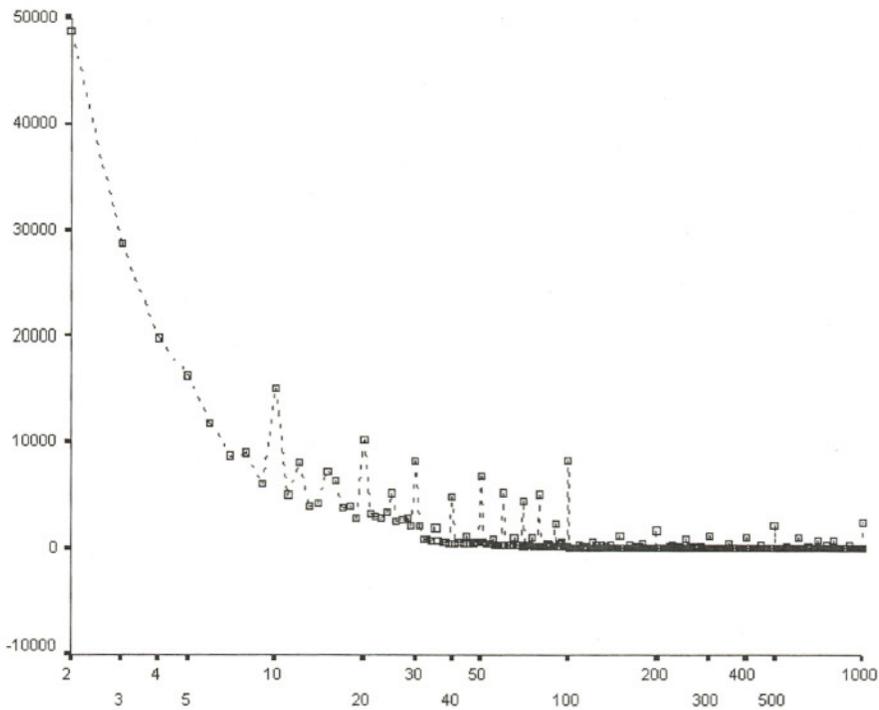


Fig. 1. Observed number frequencies.

The line depicting the number frequencies in Figure 1 has three striking characteristics:

- its slope, often related to magnitude in the numerical cognition literature. The smaller the number, the more frequently it will be used (cf. Dehaene & Mehler, 1992);
- its peaks at the numbers 10, 12, 15, 20, 30, 100, 500, etc.; these peaks are caused by what may be called a 'roundness effect';
- its partition into three sections:
  - in the first section, ending at number 9, the slope is almost perfect and peaks are hardly visible;
  - in the second section, from 10 to 99, the frequencies are substantially lower than in the first section, and now slope and peaks together determine the frequency distribution;

- in the third section, from 100 to 1000, the frequencies have again dropped considerably, and the peaks determine the picture even more than they did in the second section.

Below we will focus on the roundness issue, in relation to the effect of magnitude. The partition phenomenon will be treated in another publication (Pollmann & Jansen, in preparation).

As is clearly visible from Figure 1, the peaks of the round numbers seem to contradict the effect of magnitude. There are numbers (usually called 'round'), which clearly occur more frequently than their smaller neighbours in the row of numbers do. How can this be explained? In the light of the effect of magnitude, why would one find 'round' numbers like 10, 25 and 100 more often than numbers like 9, 23 or 97? In our opinion the solution to the conflict has to be found in a better understanding of roundness in the first place. This is what we would like to undertake in this paper.<sup>1</sup>

### THE ROUNDNESS EFFECT

Dehaene and Mehler (1992, p. 18) propose an explanation of the high frequencies of round numbers. They compared the frequency of numerals and ordinals in American English, Catalan, Dutch, French, Japanese, Kannada (a language spoken in Southern India), and Spanish. The number word frequencies were taken from standard databases of these languages. Referring to Rosch (1975), Dehaene and Mehler (1992) claim that while there are some numbers which are conventionally used for precise denotation of numerosity, there is also a group of numbers (the 'reference' numbers as Dehaene and Mehler call them) which are most suited for use in estimates. Language users can use such numbers to indicate not only the quantity represented by the number itself, but also quantities that could be more precisely described by

---

<sup>1</sup>As a possible explanation for the number frequency data we found, one might consider Benford's law. This law, of which the origin is still poorly understood (cf. Dehaene, 1997), says that the probability of a number starting with a small digit is always higher than the probability of a number starting with a higher digit:  $P(n) = \log_{10}(n+1) - \log_{10}(n)$ . As is clearly shown in Dehaene and Mehler (1992, p. 10–12), however, Benford's law has only a limited power when it comes to explaining the frequency data of numerals. It applies only to the leftmost digit in a number and not to whole arabic numbers or number words, and it does not account in any way for the frequency peaks found for numbers like 10, 25, and 100.

numbers close in value to the reference number. Therefore, it can be expected that round numbers will be used more frequently than other numbers.

This explanation, however elegant, leaves an interesting question open for discussion: what is it that makes a number round? Dehaene and Mehler apparently regard roundness as an absolute rather than as a relative number characteristic: a number is round, or it is not round. They could only speculate about how it happens that in various languages the same numbers are apparently always used as reference numbers (1992, p. 18). According to them, the reference numbers that they encountered in the language corpora they investigated, are either 'multiples of the powers of the base (base 10)' or 'integer divisors of 60'. Unfortunately, Dehaene and Mehler do not discuss why the numbers 6 and 60, which are the most prominent integer divisors of 60, play only a minor role in the frequency distributions that they present. Neither do they address the question of how it happens that multiples are important with the 'powers of base 10', and divisors with 'base 60'.

#### ROUNDNESS AND MAGNITUDE

Dehaene and Mehler did not use the frequencies of all the number types in a certain range. For all languages in their study they counted: number words from 'zero' to 'nine'; number words from 'ten' to 'nineteen'; decade number words from 'ten' to 'ninety'; number words for powers of ten ('one', 'ten', 'hundred', 'thousand', 'million', 'billion'), ordinals from 'first' to 'ninth', and ordinals from 'tenth' to 'nineteenth'. For Japanese they counted: Arabic numerals 0-9 and the decades 10-90. For French, they used the Arabic numerals 0-20, 30, 40, 50, 60, 70, 80, 90, 100 (Dehaene & Mehler, 1992, pp. 3-5).

To find a more satisfactory answer to the question of what makes a number suitable for use in an approximation context, we carried out an investigation in which we were able to work with more detailed empirical material. In our first investigation, we used the results of a frequency count in our Dutch newspaper corpus of all numbers between 2 and 1000 as data, number words, and Arabic numerals alike. We systematically kept the number 1 outside the data collection. In Dutch, as in other languages, the word for the number 1 has the same form as the indefinite article, and we saw no way to select the number word uniquely. The frequencies of numbers between approximately 1200 and 2100 are heavily influenced by the use of these numbers as calendar years (cf. Pollmann & Baayen, 2001). Therefore, we kept numbers larger than 1000

outside the data collection. We counted numbers like 'ten thousand' or '32 million' as instances of '10' and '32' respectively. Numbers such as '10,000' and '32,000,000' were not processed.

First of all, we investigated whether the frequency distribution of the numbers in our Dutch corpus can be described with factors such as those reported by Dehaene and Mehler:

- the *number*  $n$  itself (more precisely: the value of a power function, or a polynomial function of the number);
- the *reference value* of the number, or its *roundness*. Following the example of Dehaene and Mehler (1992, p. 18), we interpreted this concept as a measure of suitability for use in an approximation context.

As the operational definition of the concept 'reference value/roundness' we used the frequency of the number in a typical approximation context in the same Dutch corpus. For every number between 2 and 1000, we determined how often that number immediately followed the approximation word 'ongeveer' (English: about, approximately) in our corpus. Doing this, we assumed that reference value/roundness is not an absolute but a relative characteristic.

In the analysis of our Dutch 27-mln-corpus, we found 363,111 number tokens between 2 and 1000. In 0.43% of these occurrences (1604 cases) that number was immediately preceded by 'ongeveer'. Linear regression analysis with  $F$  (frequency) as criterion variable and  $n^{-1}$  (the number  $n$  itself raised to the power of  $-1$ ),  $n^{-2}$ ,  $n^{-3}$ , and  $n^{-4}$  as candidate predictors made it clear that the variance in  $F$  could be largely explained by a polynomial function ( $R^2 = 0.922$ ). For two of the variables involved in the analysis ( $n^{-1}$  and  $n^{-2}$ ), the contribution to the explained variance was statistically significant ( $\alpha = 0.05$ ).

In Figure 2 the relationship is shown between the frequencies that were actually observed for the numbers 2–1000 and the frequencies as predicted using the equation with  $n^{-1}$  and  $n^{-2}$  as it resulted from the regression analysis. As Figure 2 shows, the correspondence between the observed and predicted frequencies, in spite of the high  $R^2$  in the regression analysis, is low where the observed frequencies reach peak values: for the number 10, for example, but also for numbers such as 20, 50, 100, 500, and 1000: thus for numbers that are usually called 'round'. In Figure 3 the roundness of these and other numbers is presented, expressed in  $FA$ : the frequency of occurrence in an approximation context, here: after the word 'ongeveer'. Figure 3 makes it clear that roundness is indeed a relative characteristic. It is not the case that just a select part of the numbers (only the numbers 2–9 and the powers of ten,

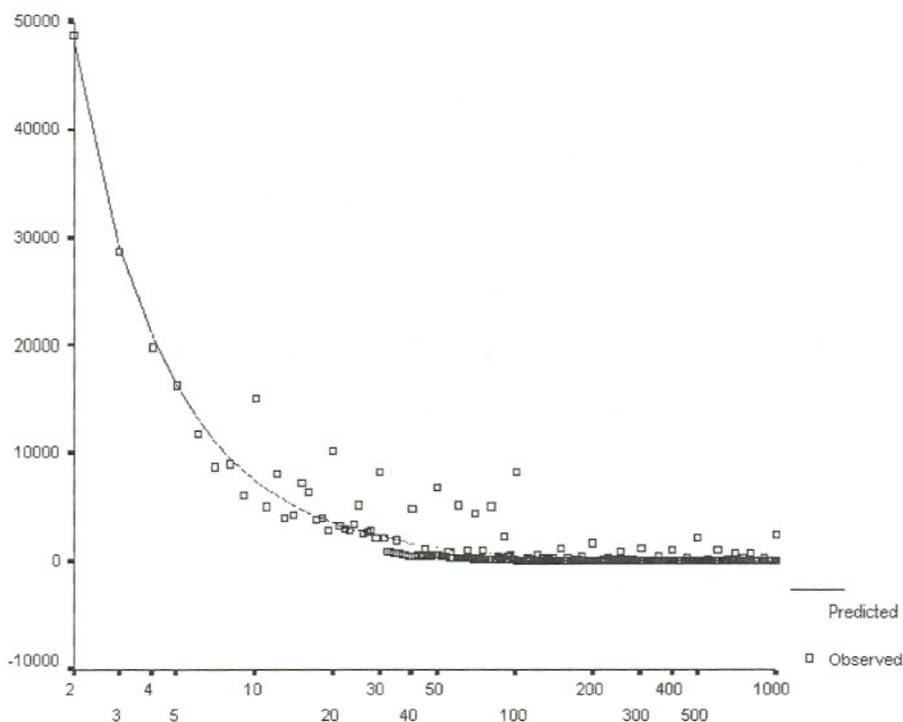


Fig. 2. Number frequency predictions (predictors: powers of number size  $n^{-1}$  and  $n^{-2}$ ).

for instance) are used in the approximation contexts. Other numbers prove to be used after an approximation expression word as well.

From a new regression analysis, again with  $F$  as criterion, but now with  $FA$  as candidate predictor in addition to  $n^{-1}$  and  $n^{-2}$ , it emerged that all of the variables contributed significantly ( $\alpha = 0.05$ ) to the explained variance in  $F$ . A very high percentage (98.7%) of the variance in number frequencies could be explained with this combination of a polynomial function and  $FA$ : the roundness values of  $n$  (see Fig. 4).

#### ROUNDNESS AS A CROSS-LINGUISTIC PHENOMENON

Now that roundness has been identified as an important predictor of frequency, the question arises as to whether this numerical characteristic is indeed, as

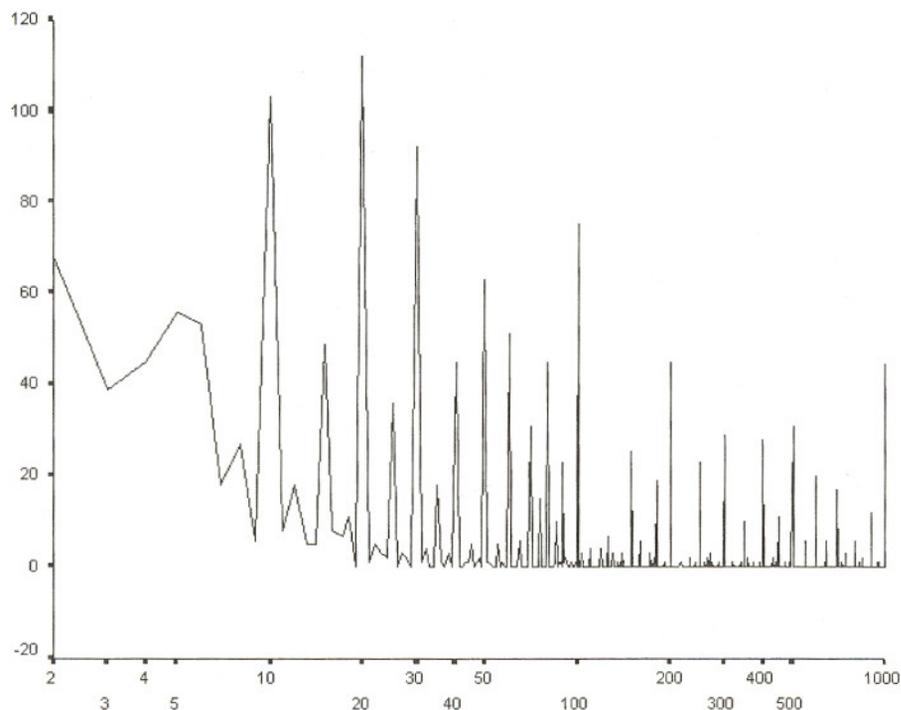


Fig. 3. FA: Frequencies of numbers in an approximation context.

Dehaene and Mehler (1992, p. 18) suggested, a general, cross-linguistic phenomenon. In order to gain insight into this matter, we added comparable data from three other languages (German, French, and English) to the roundness data we had collected for Dutch. For German, French, and English, we made use of CD-ROMs that contained texts which had originally appeared in the 1994 volumes of three major newspapers published in Germany, France, and England: *Die Frankfurter Allgemeine Zeitung* (about 27 million words), *Le Monde* (about 26 million words), and *The Times* (about 40 million words), respectively. We used as approximation contexts [German 'etwa', directly followed by a number], [French 'environ', directly followed by a number], and [English 'about', directly followed by a number], and just as we did for the Dutch corpus, we determined the frequencies of the numbers 2 through 1000 in these approximation contexts.

For German we found 9017 instances of numbers in approximation contexts; for French, 1340; and for English, 6825. The correlation coefficients

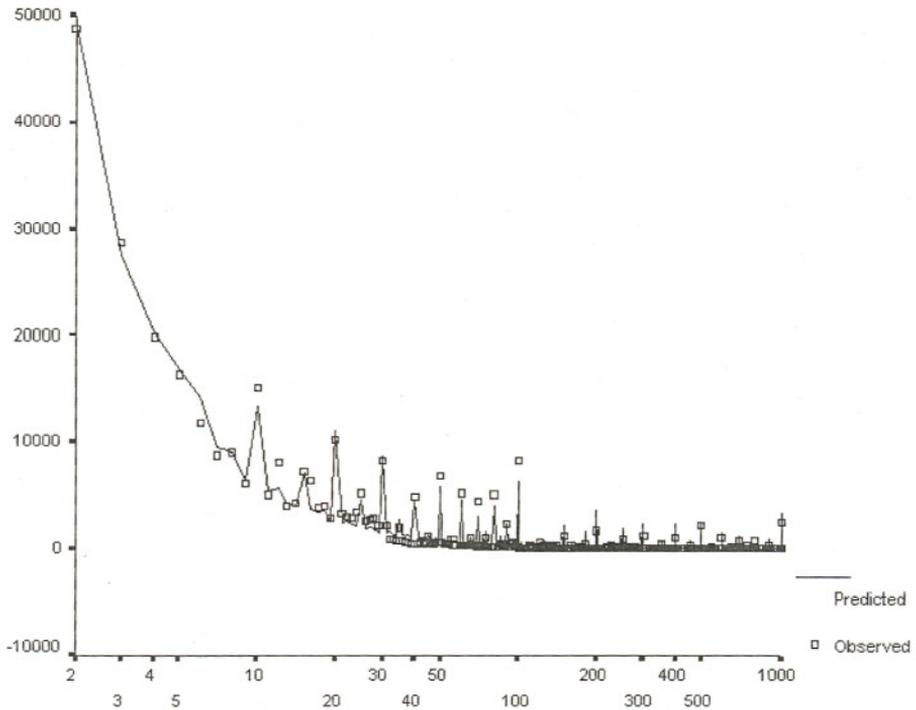


Fig. 4. Number frequency predictions (predictors: powers of number size ( $n^{-1}$  and  $n^{-2}$ ) and *FA*: frequency of occurrence in an approximation context).

are shown in Table 1. The idea of Dehaene and Mehler (1992, p. 18) that the roundness of a given number does not differ substantially per language or culture, but that it is a cross-cultural constant, is clearly reflected by the results given in Table 1. Not a single correlation coefficient is lower than 0.885.

#### DISTINCTIVE FEATURES OF ROUNDNESS

The problem is still unsolved as to which factors determine how round a number is. In the literature, this question has only been posed by Dehaene and Mehler (1992), Rosch (1975), and Sigurd (1988). Earlier we indicated why the answers of Dehaene and Mehler (1992) and Rosch (1975) are not completely satisfying. Sigurd (1988) takes the interesting position that round numbers

Table 1. Correlations Between Frequencies in Approximation Contexts in Newspaper Corpora in Four Languages.

	Dutch	German	French	English
Dutch	1.000			
German	0.970	1.000		
French	0.887	0.926	1.000	
English	0.947	0.959	0.885	1.000

typically are related to the base of the number system that is used. In a base-10 system, Sigurd considers as base numbers the powers of 10, beginning at 10 itself: 10, 100, 1000, 10,000, etc. Based on the idea that in a decimal system not only the powers of ten are rounder than other numbers, but that the divisors ('simple fractions') of the powers of ten are also rounder, Sigurd presents an equation where the roundness of a number is determined by its divisibility by a base number, half of a base number, and a quarter of a base number.

However thought-provoking Sigurd's idea may be, we are not entirely convinced by his approach to the roundness issue. The theoretical foundation is essentially the intuition that roundness has to do with the 'mathematical properties (factorial contents) of the numbers' (Sigurd, 1988, p. 249). It remains unclear, however, why the half and the quarter of a base number are the numbers that contribute to the roundness, and not, for example, one-fifth of the base (2, 20, 200, etc.).

Sigurd's attempts to demonstrate the value of his equation are not really convincing either. Referring to frequency distributions of numbers and numerals (1-100) in Swedish and English language corpora (respectively Allén, 1970; Kučera & Francis, 1967), Sigurd states that 'the derived roundness values display clear co-variation with the frequencies of the numbers' (Sigurd, 1988, p. 250). More detailed information in the form of, for example, correlation coefficients, is absent.

We attempted to find out what it is that makes a number round by taking two steps. First, we made use of the results of an earlier study we carried out into characteristics of so-called two-number approximative expressions. In our study, we analysed pairs of numbers such as they occur in expressions like: [about 5 or 6 books], [about 10 or 12 articles], [about 20 or 25 papers]. Clearly not every combination of numbers can be used in such expressions: compare, for instance, [\* about 10 or 13 grandparents], [\* about 80 or 91 grandchildren],

or [\* about 350 or 300 presents].<sup>2</sup> Below, we will show that these expressions in French, German, or English can all be reduced to elementary numerical characteristics in much the same way as proved to be the case in Dutch. Subsequently, we will show that these elementary numerical characteristics are the same as those that determine the roundness of an individual number.

In our study (Pollmann and Jansen, 1996) we showed which number combinations in Dutch can be used in two-number approximative expressions. The members of these expressions were found to (nearly always) follow each other directly in an arithmetic sequence in which the ratio and first number are equal to:

- $1 * 10^n$ , or
  - $2 * 10^n$ , or
  - $1/2 * 10^n$ , or
  - $1/4 * 10^n$  ( $n$  always  $\geq 0$ )
- (further: the *sequence rulé*)

To give some examples of approximation pairs which would be in agreement with this *sequence rule*: [3, 4] comes from the arithmetic sequence 1, 2, 3, 4, [...] with ratio and first number  $10^0$ ; [40, 50] from the sequence 10, 20, 30, 40 [...] with ratio and first number  $10^1$ ; [18, 20] from the sequence 2, 4, 6, 8 [...] with ratio and first number  $2 * 10^0$ ; [100, 150] from the sequence 50, 100, 150, 200 [...] with ratio and first number  $1/2 * 10^2$ ; and [100, 125] from the sequence 25, 50, 75, 100 [...] with ratio and first number  $1/4 * 10^2$ . Combinations like [\*1, 3], [\*5, 7], [\*6, 9], or [\*40, 80], which would form part of sequences which do not correspond with the sequence rule, did not occur in the Dutch text corpus analysed in Pollmann and Jansen (1996).

Does the sequence rule hold only in Dutch for the construction of two-number approximative expressions, or are the characteristic properties of these pairs cross-linguistic? To be able to answer that question, we also analysed two-number approximative expressions in German, French, and English.

In our four language corpora we looked for pairs of numbers which could be interpreted as approximative expressions of quantities. In the French corpus we looked for combinations of two numbers coordinated by [à]. For English we confined the data to combinations of numbers in coordinations with [or]. In

<sup>2</sup>Sigurd (1988, pp. 245–247) also pays attention to two-number approximative expressions. The regularities he brings to the fore diverge to some extent from our findings. It should be noted, though, that his statements about these expressions are solely based on his own intuitions and not on frequency data in a language corpus.

German as in Dutch, such a conjunction is not necessary to express an approximation. It turned out that the sequence rule, originally based on Dutch data, is consistent with data from German, French, and English. In none of the four languages were more than 10% of the two-number approximative expressions that we found in contrast with the sequence rule. In Dutch 90.8% of the two-number approximative expressions turned out to be in accordance with this rule, in German 93.8%, in French 93.3%, and in English 98.2%.

We conclude that cross-linguistic regularities do exist in the appearance of two-number approximative expressions, and that these regularities can be described by a sequence rule where a small set of basic means to manipulate quantities plays a central role. It was apparent that the predictive power of the sequence rule could be attributed, to a high degree, to the frequent occurrence of expressions where first number and ratio equal a base number of the decimal system  $1 * 10^n$  (between 63.0 and 91.7% of the cases). There was also a relatively large contribution in all four languages of expressions where first number and ratio equal  $2 * 10^n$  or  $1/2 * 10^n$  (between 2.1 and 9.0%). In 4.3 to 21.5% of the cases we found expressions where first number and ratio equal  $5 * 10^n$ . Expressions where first number and ratio equal  $1/4 * 10^n$  hardly proved to be used, however. In no language did we find more than 0.5% of these expressions. Therefore, in contrast to what we claimed in our study (Pollmann and Jansen, 1996), there seems to be no place in this rule for sequences based on  $(1/4 * 10^n)$  as first number and ratio. Our rule, therefore, only comprises sequences based on

- $(1 * 10^n)$
  - $(2 * 10^n)$
  - $(1/2 * 10^n)$  ( $n$  always  $\geq 0$ )
- (further: the *revised sequence rule*)

The quarters almost being absent as a determining factor in approximation pairs sheds some doubt on the correctness of Sigurd's conjecture that the quarter of a number is a determining factor in roundness at all. We return to this issue below.

#### FROM PAIRS TO SINGLE NUMBERS

An obvious idea is that from these three types of sequence not only pairs can be drawn that are suited for use in an approximation context, but also numbers

that can be used alone after an approximation word. In order to test this idea, we constructed three dichotomous variables:

- 10-ness: 'is the number contained in the set  $[1 * (1-9 * 10^n)]$ ?'
- 2-ness: 'is the number contained in the set  $[2 * (1-9 * 10^n)]$ ?'
- 5-ness: 'is the number contained in the set  $[5 * (1-9 * 10^n)]$ ?' ( $n$  always  $\geq 0$ )

Formulated somewhat less formally: a number has '10-ness' when the result of dividing it by 1, 10, 100, 1000, etc. is 1, 2, 3, 4, 5, 6, 7, 8, or 9; it has '2-ness', when dividing it by 2, 20, 200, 2000, etc. results in 1, 2, 3, 4, 5, 6, 7, 8, or 9; and it has '5-ness' when one of these outcomes is reached by dividing it by 5, 50, 500, 5000, etc. To give some examples: 40 has 10-ness, 2-ness, and 5-ness, 8 has 10-ness and 2-ness but no 5-ness, 300 has 10-ness and 5-ness but no 2-ness, 70 has only 10-ness, and 61 has none of these properties.

How well can the variance in *FA* (frequency in an approximation context) be predicted using these variables 10-ness, 2-ness, and 5-ness together? Are numbers which belong to more than one of the sets defined by 10-ness, 2-ness, and 5-ness, more suited for use in an approximation context than other numbers? Is the number 20, for instance, which belongs to all three sets, really 'rounder' than the number 80 (10-ness and 2-ness)? To give another example: is the *FA*-value of the number 50, with its 10-ness and 5-ness, higher than the *FA*-value of the number 45 (only 5-ness)?

A new linear regression analysis, now with *FA* as criterion variable and 10-ness, 2-ness, and 5-ness as candidate predictors showed that 79.6% of the variance in *FA* was explained by the combination of the variables involved, and that each of these contributed significantly ( $\alpha = 0.05$ ). It may be concluded, therefore, that 10-ness, 2-ness, and 5-ness predict, to a reasonably large extent, the roundness of a number: its frequency in approximation contexts.

## NUMBERS IN ARBITRARY CONTEXTS

Given the obvious role that roundness plays in the frequency of numbers in arbitrary contexts, the suggestion seems natural that the combination of 10-ness, 2-ness, and 5-ness of a number also determines to a substantial extent how often that number will be used *irrespective of its context*. That idea was confirmed by another regression analysis, now with *F* as criterion variable and  $n^{-1}$ ,  $n^{-2}$ ,  $n^{-3}$ ,  $n^{-4}$ , 10-ness, 2-ness, and 5-ness as candidate predictors. Together, the variables  $n^{-1}$ ,  $n^{-2}$ , 10-ness, 2-ness, and 5-ness proved to explain

96.6% of the variance in  $F$ . Each of these predictors contributed significantly; the addition of  $n^{-3}$  or  $n^{-4}$  to the regression equation did not lead to a significant increase in the explained variance ( $\alpha = 0.05$ ). These results demonstrate that it is very well possible to predict the variance in number frequencies, using a few basic numerical properties: a polynomial function of the magnitude of the number, and the characteristics 10-ness, 2-ness, and 5-ness.

We also computed the contribution to number frequency of other kinds of possibly relevant factors, like 3-ness, 4-ness, 6-ness, 7-ness, etc. None of these turned out to contribute significantly.

Although – as reported above – the outcomes of our analysis of approximation pairs based on the factor ‘quarter of the base’ emerged not to be promising, we nevertheless calculated the contribution of ‘2½-ness’ to number frequency, ‘2½-ness’ being defined as a positive answer to the question: ‘is the number contained in the set  $[2\frac{1}{2} * (1-9 * 10^n)]$ ?’

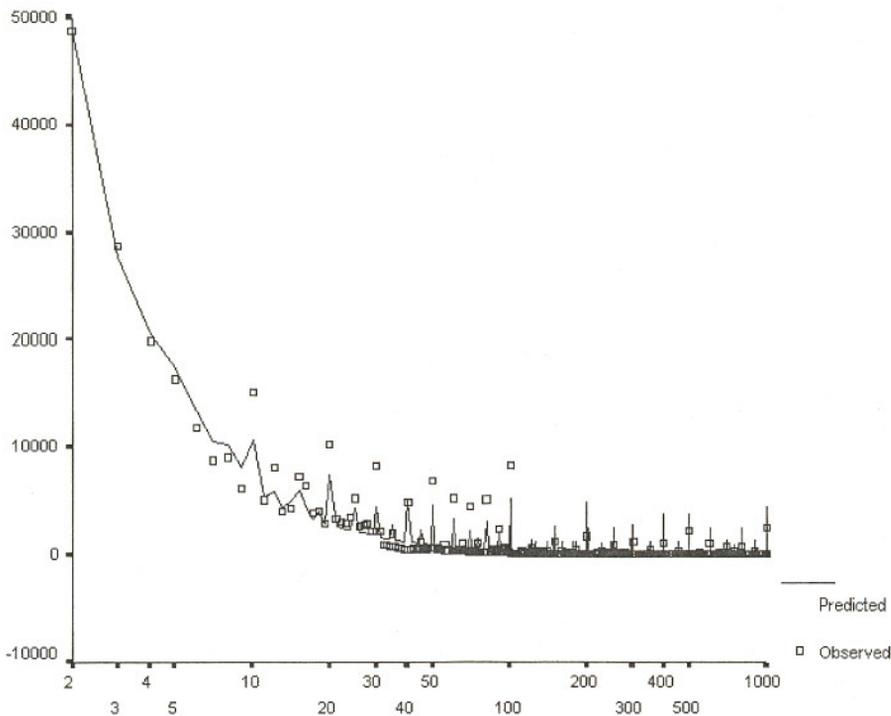


Fig. 5. Number frequency predictions (predictors: powers of number size ( $n^{-1}$  and  $n^{-2}$ ), and 10-ness, 2-ness, 5-ness, and 2½-ness).

This time the computations show a small, but significant contribution of the 'quarters of the base' factor. The resulting equation formula is:

$$nf = 52562 * nm^{-1} + 79933 * nm^{-2} + 1467 * te + 1044 * tw \\ + 1266 * fi + 1009 * qu - 135$$

where  $nf$  is 'number frequency',  $nm$  is 'number magnitude',  $te$  is '10-ness',  $tw$  is 2-ness,  $fi$  is 5-ness, and  $qu$  is  $2\frac{1}{2}$ -ness.  $R^2$  (total proportion of variance explained): 0,968.

We therefore claim to have found an empirical and statistical underpinning of Sigurd's ideas, and additionally to have shown that 2-ness is a numerical property that contributes significantly to roundness of numbers, and to the frequency with which they are used in text.

Figure 5 visualizes the strong relationship between the frequencies observed for the numbers 2–1000 and the frequencies as predicted using  $n^{-1}$ ,  $n^{-2}$ , 10-ness, 2-ness, 5-ness, and  $2\frac{1}{2}$ -ness.

## CONCLUSION

To summarize our findings: in cultures with a decimal number system, the degree of roundness is closely associated with four numerical properties: 10-ness, 2-ness, 5-ness, and  $2\frac{1}{2}$ -ness. The more of these features a number shares, the rounder it is, and the higher its frequency.

As an explanation for this apparent favouritism of numerical expressions with 10-ness, 2-ness, 5-ness, and  $2\frac{1}{2}$ -ness, it seems reasonable to assume that doubling and halving (sometimes followed by halving again) are basic means to manipulate quantities. Apart from its influence on the frequency of numbers in text corpora, we can see this *principle of favourite quantities* at work – at least in western societies – in, for instance, the denominations of coins and notes (to give an example, from 2002 on we have coins of 1 euro, 2 euro, 5 euro, 10 euro, 20 euro etc.), and in, for instance, the years when we celebrate our jubilees (5, 10, 25, 50, etc.).

To what extent the principle of favourite quantities also works in duodecimal or sexagesimal number systems is hard to determine, as these systems exist only in very rudimentary forms today. But what we have left, especially in the field of time measuring, seems to support the favouritism of halving, doubling, and sometimes ~~halving~~ again. We partition our time periods into

~~Halving~~  
H halving

two halves and four quarters. An academic year, for instance, is divided into two semesters and four trimesters, and the traditional way to design watch faces is to segment them into two periods of 30 min (or 6 hr) and four periods of 15 min (or 3 hr). Perhaps it is precisely this effect of the principle of favourite quantities in a duodecimal system that led Dehaene and Mehler (1992) to attribute a special status to the number 60 and its integer divisors – a status that is not substantiated by our frequency data, however.

Apparently, there is a 'natural' propensity for halving and doubling quantities. This propensity – be it innate or 'just easily learned' – seems to be independent from the counting system that is being used (see also Pollmann & Jansen, 1996, p. 234–236).

#### ACKNOWLEDGEMENTS

We wish to thank H. van den Bergh and E. Savelsbergh for their comments and advice.

#### REFERENCES

- Allén, S. (1970). *Nusvensk frekvensordbok 1*. Stockholm: Almqvist & Wicksell.
- Dehaene, S. (1997). *The number sense. How the mind creates mathematics*. London: Penguin Press.
- Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1–29.
- Kučera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.
- Pollmann, T., & Baayen, H. (2001). Computing historical consciousness. A quantitative inquiry into the presence of the past in newspaper texts. *Computers and the Humanities*, 35, 237–253.
- Pollmann, T., & Jansen, C. (in preparation). *Scales as a determining factor in number frequency in texts*.
- Pollmann, T., & Jansen, C. (1996). The language user as an arithmetician. *Cognition*, 59, 219–237.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7, 532–547.
- Sigurd, B. (1988). Round numbers. *Language in Society*, 17, 243–252.