# "FOUR-AND-TWENTY, TWENTY-FOUR". WHAT'S IN A NUMBER?

Knut Kvale<sup>1</sup> and Arne Kjell Foldvik<sup>2</sup>

1. Telenor Research and Development, N-2007 Kjeller, Norway 2. Dept. of Linguistics, Norwegian University of Science and Technology, N-7055 Dragvoll, Norway

## ABSTRACT

This paper investigates "what's in a number", i.e. how natural numbers are pronounced in several European languages. As regards reading numbers above 20, 29 languages read the decade first and then the digit, e.g. twenty-four, and 10 languages read the digit before the decade, e.g. four-and-twenty. Two languages, Norwegian and Czech, use both systems, and 9 languages use (partly) a vigesimal system.

An analysis of the Norwegian part of the European SpeechDat database showed that reading the decade first is used more in formal than in non-formal (spontaneous) speech and that typographic layout of digits influenced the reading of them.

## **1. WHAT'S A NUMBER?**

A decimal number such as 7392 represents a quantity equal to 7 thousands plus 3 hundreds, plus 9 tens, plus 2 units. The thousands, hundreds, etc., are powers of 10 implied by the position of the coefficients. To be more exact, 7392 should be written as:  $(7x10^3) + (3x10^2) +$  $(9x10^1) + (2x10^0)$ . However, the *convention* is to write only the coefficients and from their *position* deduce the necessary powers of 10. The decimal number system is said to be of base, or radix, 10 because it uses ten digits and the coefficients are multiplied by powers of ten. The binary system, on the other hand, is of base 2, i.e. it has only two digits, 0 and 1, which are multiplied by powers of 2. The decimal equivalent of the binary number 101 is therefore  $(1x2^2) + (0x2^1) + (1x2^0) = 5$ .

# 2. NUMBERS IN EUROPEAN LANGUAGES

Even if the European languages agree on the quantity of a number such as 7392, they show up different counting systems and ways of reading numbers. As regards reading numbers above 20, an important difference between languages in Europe is the ordering of the decade and the digit. By far the largest group of languages read the decade first and then the digit e.g. twenty-four, while other languages read the digit first, e.g. four-and-twenty. This phenomenon may be due to the fact that addition is commutative, i.e. 20+4 = 4+20.

As an example of the diversity in number reading, table 1 shows the pronunciation of 21 in several European languages.

Language	Pronunciation of 21	English equivalent
Basque	hogeita bat	(one) score one
Bosnian	dvadeset jedan	twenty one
Croatian	dvadeset jedan	twenty one
Czech	dvacet jedna	twenty one
English	twenty-one	twenty-one
Estonian	kakskümmend üks	twenty one
Finnish	kaksikymmentäyksi	twentyone
Greek	['ikosi 'ena]	twenty one
Italian	ventuno	twentyone
Lithuanian	dvides^imt vienas	twenty one
Norwegian	tjueen / tjueein	twentyone
Polish	dwadzies'cia jeden	twenty one
Russian	dvadsat odin	twenty one
Sami	guoktelogiokta	twentyone
Serbian	dvadeset jedan	twenty one
Slovak	dvadsat^ jeden	twenty one
Spanish	ventiuno	twentyone
Swedish	tjugoett	twentyone
Turkish	yirmibir	twentyone
Ukranian	dvadtsyat' odyn	twenty one
Albanian	njëzet e një	twenty and one
Bulgarian	dvadeset i edno	twenty and one
Catalan	vint-i-un	twenty-and-one
French	vingt-et-un	twenty-and-one
Hungarian	huszonegy	twentyandone
Icelandic	tuttugu og einn	twenty and one
Irish	fiche is a-haon	twenty and one
Portugese	vinte e um	twenty and one
Romanian	douazeci si unu	twenty and one
Breton	unan warn-ugent	one on twenty
Czech	jednadvacet	onetwenty
Danish	en og tyve	one and twenty
Dutch	eenentwintig	oneandtwenty
German	einundzwanzig	oneandtwenty
Norwegian	enogtjue / einogtjue	oneandtwenty
Scottish	aon air fhichead	one and twenty
Gaelic		
Slovenian	enaindvajset	oneandtwenty
Welsh	un ar hugain	one on twenty

# **Table 1** Pronunciation of the number 21 in 36European languages

European languages that read the decade first and then the digit are: Albanian, Basque, Bosnian, Bulgarian, Catalan, Croatian, Czech, English, Estonian, Finnish, French, Greek, Hungarian, Icelandic, Italian, Lettish, Lithuanian, Norwegian, Polish, Portuguese, Romanian, Russian, Sami, Serbian, Slovak, Spanish, Swedish, Turkish and Ukrainian. Of these languages the following insert *and* between the decade and the digit: Albanian, Bulgarian, Catalan, French (only with the digit one), Hungarian, Icelandic, Irish, Portuguese, Romanian.

European languages that read the digit before the decade include: Breton, Czech, Danish, Dutch, Frisian, German, Norwegian, Scottish Gaelic, Slovenian and Welsh. All of them link digit and decade with *-and -*, one-and-twenty, except for Czech which has no linking conjunction, and Breton and Welsh which use *on* instead of *and*.

As for the decades, most of the languages name the decades in the same way as e.g. English does. But in some of the languages counting of the decades is based on the number twenty. Languages that at least partly use this vigesimal system are: Albanian, Basque, Breton, Danish, Faroese, French, Irish, Scottish Gaelic and Welsh.

Earlier both ways of counting, digit before the decade and decade before the digit, were used in Germanic languages. For instance in Old Norse 21 was written both as "einn ok tottogo" and "tottogo ok einn" [1]. However, reading the digit before the decade was most common and gradually took over completely in Norwegian, probably because this pronunciation was more in accordance with the normal stress pattern in the language [2]. Also in Danish and Faroese digit before the decade took over, as in German and Dutch, while in Swedish, Icelandic and English decade before the digit replaced digit before the decade.

Basque, Hungarian and Turkish read the decade before digit even with the numbers 11 to 19. Several other languages do this for some of these numbers, e.g. French, Italian, Portuguese, Spanish and Welsh.

# 3. COUNTING IN NORWEGIAN

In Norwegian counting the digit before the decade as well as the decade before the digit is used. The background for this unusual situation in Norwegian is that it was decided through an act in parliament that from July 1, 1951 numbers above 20 should be read with the decade before the digit, sixty-four, and not as previously, four-and-sixty [2]. One of the arguments for the reform was that it is more *logical* to read the numbers in the order they appear in the text. Table 1 shows, however, that one way of number reading is not more logical than another - a language is only a set of conventions that people have got used to.

Norwegian, unlike languages such as English, has had fairly frequent spelling reforms during the last hundred years. The aim has been to keep the spelling close to the pronunciation. The 1951-reform was, however, quite unique in that it changed not only the orthography of a very frequently used group of words, but it also changed the word order, the syntax (no co-ordinator), and the pronunciation (both phonemes and rhythm) [2].

### 3.1 The Norwegian SpeechDat database

We have analysed the pronunciation of numbers in the Norwegian part of the European SpeechDat database [3,4] which contains 1016 speakers from all over Norway recorded over real telephone lines in 1997. The country was divided into 23 dialect regions and the number of informants is representative with respect to dialect and sex. People were contacted by mail and given a description of the project and a manuscript. About 20 % agreed to take part and phoned up an automatic service which prompted them to read sentences from the manuscript.

Each informant read among other items an eight digit telephone number, a sixteen digit credit card number, a sequence of ten digits (with pauses), an eight digit prompt sheet number, a six digit PIN-code, two natural numbers, and a currency amount. The service also asked two questions: "When were you born?" and "What's the time?", which were spontaneously answered.

In the analysis we have regarded all words as properly pronounced although some words were marked as mispronounced or the signal was truncated in some part of the word. In both cases the words were perceived as intelligible by the annotators.

## 3.2 Use of new and old counting in Norwegian

Table 2 summarises the use of old counting (decade before digit) in the Norwegian SpeechDat database with respect to the informants' age. Typically, in non-formal speech and expressions such as date of birth, the old pronunciation is more common than in formal speech such as reading telephone numbers. However, of the 43.2% (439 speakers) who used the old counting at least once, about half of them (53 %) used the old counting only once, while 25 % used it twice. That is, most of them mixed the two counting styles.

Bearing in mind that the reform in number reading was carried out nearly 50 years ago, the differences between the age groups 19-34, 35-59 and the 60+ in the use of the old way of counting for date of birth are surprisingly small.

Norwegian has two slightly different orthographies called *bokmål* and *nynorsk*. Each municipality chooses which orthography to use in their administration. Manuscripts were sent to the informants in accordance with the official orthography for their municipality, giving 880 (86.6 %) informants reading from a bokmål text and 136 (13.4%) from a nynorsk text.

The informants used the old counting equally much regardless of bokmål and nynorsk manuscripts. (Bokmål 43.1 %, nynorsk 44.1%).

Age	No. of informants	Percentage	Old in read phone numbers	Old in read credit cards	Old in date of birth (spontaneous)	Old in time of day (spontaneous)	Total use of old in numbers
-18	9	0.9 %	-	-	-	-	-
19-34	418	41.1 %	2.2 %	1.4 %	28.7 %	3.3 %	44.7 %
35-59	409	40.3 %	2.7 %	2.4 %	24.0 %	5.9 %	39.6 %
60+	142	14.0 %	10.6 %	17.6 %	31.0 %	6.3 %	50.0 %
Unknown	38	3.7 %	2.6 %	5.3 %	34.2 %	2.6 %	50.0 %
Total	1016	100 %	3.5 %	4.2 %	27.1 %	4.7 %	43.2 %

Table 2Number of informants using the old counting (decade before digit) at least once.The right column sums up all the number sequences listed in section 3.1

#### 3.3 The numbers 7, 20, 30 and 40

The 1951-reform also changed the *pronunciation* and *spelling* of the numbers 7, 20, 30 and 40 (the forms *syv*, *tyve*, *tredve* and *førr* were substituted by *sju*, *tjue*, *tretti* and *førti* respectively). Table 3 shows that the *new* forms for 20, 30 and 40 are more frequently used in read numbers than in spontaneous speech. The use of sju and syv may be considered as normal dialect variants, not as an indication of formal or informal speech.

Number string	sju	tjue	tretti	førti
Telephone numbers	71 %	96 %	95 %	98 %
Credit card numbers	60 %	92 %	90 %	99 %
Date of birth	62 %	80 %	71 %	83 %
Time of day	73 %	81 %	62 %	95 %
Total	60 %	89 %	84 %	96 %

**Table 3** Percentage occurrencies of the 'new' forms sju, tjue, tretti and førti of 7, 20, 30, and 40 respectively. The total summarises the occurrences of new forms in all the number sequences in the Norwegian SpeechDat database (see section 3.1)

Table 4 shows that informants who use the old forms of 7, 20, 30 and 40 are more likely to use the old counting (decade before digit) than informants who use the new forms of these numbers.

	New forms of 7, 20, 30,	Old forms of 7, 20, 30,
	40 and old counting	40 and old counting
7	288/732 = 39.3 %	261/513 = 50.9 %
20	323/815 = 39.6 %	107/140 = 76.4 %
30	273/686 = 39.8 %	113/177 = 63.8 %
40	286/689 = 41.5 %	33 / 37 = 89.2 %
Total	1170/2922 = 40.0%	514/867 = 59.3 %

**Table 4** *Left column: Informants who used new forms of* 7, 20, 30 and 40 and the old counting divided by the number of informants using the new forms.

Right column: Informants who used old forms of 7, 20, 30 and 40 and the old counting divided by the number of informants using the old forms The percentages of the informants that *mixed* the old and new forms 7, 20, 30 and 40 were: 24.4 % for 7, 10.3 % for 20, 15.3 % for 30, but only 3.3 % for 40.

#### **3.4 Telephone numbers**

Norwegian telephone numbers consist of 8 digits. In the telephone directories ordinary phone numbers start with 2-7 and are written as 4 pairs of numbers, whereas e.g. mobile telephone numbers and special rate numbers start with 8 or 9 and are written as xxx xx xxx (denoted 3-2-3 numbers). In our database there are 762 4-pair numbers and 254 3-2-3-numbers.

Table 5 shows that a difference in typographic layout of digits influences the reading of them. When grouped as 4 pairs of numbers 70 % read number-pairs. 26 % mixed number-pairs and digits because one or more number-pair started with 0.

The 3-2-3 grouping implied that half the informants read the 3-digit group as hundred and something. One out of five read these numbers digit by digit. Here the mixing of number-pairs and digits was partly because one or more number-pair started with 0, and partly because people re-grouped the number, e.g. digit + number pair + number pair + digit + number pair.

	4 pairs	3-2-3
8 digits	21/762 = 2.8 %	55/254 = 21.7 %
4 number-pairs	535/762 = 70.2 %	4/254 = 1.6 %
Mixing digits and	198/762 = 26.0 %	68/254 = 27.2 %
number-pairs		
With 'hundred'	1/762 = 0.1 %	126/254 = 49.1 %
Others	6/762 = 0.8 %	1/254 = 0.4 %

**Table 5** Number of informants who used the differentgroupings of 8 digits phone numbers

The use of old counting is less in SpeechDat (table 6) than in another Norwegian database "TABU.0" [5]. In TABU.0 7.8 % of the informants used old counting when reading telephone numbers grouped as 4 number-pairs [2]. In the SpeechDat a fully automatic service gave the information and carried out the recording. This situation may have been experienced as more formal

than the TABU.0 recording session where a person phoned up the informant, gave instructions before the recording and feedback during the recording session.

Age	Old in 4 pairs	Old in 3-2-3
-18	-	-
19-34	6/327 = 1.8 %	3/91 = 3.3 %
35-59	9/289 = 3.1 %	2/120 = 1.7 %
60+	10/109 = 9.2 %	5 /33 = 15.2 %
Unknown	1/30 = 3.3 %	-
Total	26/762 = 3.4 %	10/254 = 3.9 %

**Table 6** Number of informants who used the old counting (decade before digit) at least once when reading telephone numbers, divided by the number of informants in each age-group

### 3.4 Credit card numbers

The sixteen digit credit card numbers were divided into groups of four digits. Obviously reading these numbers with 'thousand' and/or 'hundred' is less likely than when reading e.g. a sum of money, since a credit card number is experienced as a string of numbers and not as an amount. Since people want to express a number string with as little effort as possible, i.e. with as few words or syllables as possible, the grouping into number-pairs, 'twelve thirty-four', or digits, 'one two three four', are preferred to the long form 'one thousand two hundred and thirty four'.

In table 7 the group "mixing digits and number-pairs" also contains numbers which are less than 16 digits or contains other words than numbers.

Most of the 56 informants who said 'thousand' and/or 'hundred' (the three bottom rows of table 7) used this grouping consistently, giving a total of 139 'thousand' and 163 'hundred'.

The pronunciation of the numbers 1100-1999 may start both with 'one thousand' and 'x hundred'. For the credit card numbers 14 informants said 'one thousand' and 4 informants said 'x hundred and'.

Grouping	No. of informants
16 digits	336/1016 = 33.1 %
8 number-pairs	221/1016 = 20.8 %
Mixing digits and number-pairs	613/1016 = 60.3 %
With 'hundred' and 'thousand'	45/1016 = 4.4 %
With 'thousand' only	2/1016 = 0.2 %
With 'hundred' only	9/1016 = 0.9 %

**Table 7** Number of informants who used the different grouping of credit card numbers

#### CONCLUSION

A language is a set of conventions which is not necessarily always 'logical'. We have exemplified this by an overview of how natural numbers are pronounced in several European languages. For numbers above 20, 29 languages read the decade first and then the digit, e.g. twenty-four, and 10 languages read the digit before the decade, e.g. four-and-twenty. Two languages, Norwegian and Czech, use both systems.

In the Norwegian part of the European SpeechDat database, reading the decade first is more frequently used in formal than in non-formal (spontaneous) speech. As many as 43 % of the speakers read some numbers with the decade before the digit although this way of reading numbers was officially abandoned in state and municipal administration, in broadcasting and education nearly fifty years ago.

The typographic layout of digits influenced the reading of them. When a 8 digit number was grouped as 4 number-pairs 70 % read number-pairs and only 3 % read digit by digit. A 3-2-3 grouping implied that half the informants read the 3-digit group as hundred and something and that one out of five read digit by digit.

Obviously from an ASR point of view digit by digit reading of numbers facilitates the recognition task and is therefore preferable. But our Norwegian data shows that number reading habits are hard to change.

#### REFERENCES

- 1. Seip, D.A. 1950. *Tellingsmåten i norsk og mulig endring*, Land og kirke
- Kvale, K., Foldvik, A.K., 1997. "Old customs die hard: The 'new' way of counting in Norwegian is still new", Proc. FONETIKA-97, Umeå, May, 1997.
- Johansen, F.T., Amdal, I., Kvale, K. 1997. "The Norwegian part of SpeechDat: A European Speech Database for Creation of Voice Driven Teleservices", Proc. NORSIG-97, Tromsø, May, 1997.
- 4. Höge, H. et. al. 1997. "European speech databases for telephone applications", Proc. ICASSP-97, pp. 1771-1774.
- 5. Amdal, I., Ljøen, H. 1995. *TABU.0 en norsk telefontale-database*. Scientific Report 40/95, Telenor.