

Lund University, Dept. of Linguistics
Working Papers 33 (1988), 173-191

How Dates Are Made

Magnus Olsson

Abstract

The main theme in this paper is the origin and scope of certain exceptions to the universal tendency in numerals to contain their respective highest available bases. I restrict on the whole my attention to dates and periods of dates. The Swedish case is examined in some detail. This is because there has been conflicting suggestions and a quite inflamed debate as to the scope of these exceptional numerals in dates and periods of dates. I argue, contrary to the official view, that these exceptions are most naturally treated as subject to constraints at certain points in the sequence of numbers. Some other Indo-European languages are investigated for comparison. The English system exhibits unusual traits, which I give an account of in a separate section. This paper is an outgrowth of Olsson 1988 – section 4 is completely new, while 2 and 3 remain basically the same.

1 INTRODUCTION

In order to pronounce a number, languages tend to use the highest possible numeral base with its maximal high numeral attribute and then the rest of the number goes through the same procedure. Hurford 1975 (e.g. 67-80) calls this phenomenon the packing strategy and regards it (e.g. 1987:252 ff.) as being the result of social factors, while he does not think that it is a general constraint on numerals. In some Indo-European languages there is a certain way in which numerals violate the packing strategy. In English numbers with four digits, *hundred* may replace *thousand* and corresponding cases are found in e.g. Swedish.¹ Sigurd 1972:7 says that of e.g. *thirteen hundred* and *one thousand and three hundred* only the latter is maximally fused, the highest possible base being employed. In Swedish the non-maximal forms are used optionally with e.g. prices (ibid.:7) and number of employees (Olsson 1988). It is also used with dates in at least certain cases and then obligatorily. These cases are the years 1100-1999 and the centuries to which these dates belong. In the case of the centuries, a peculiar Nordic construction is used: the base *hundra* is with its attribute put before *talet* 'the number' – e.g. *fjortenhundratalet* means 'the fifteenth century'.² This con-

¹ The Swedish numeral forms that are of interest in this paper are the following. 1-9: *ett, två, tre, fyra, fem, sex, sju, åtta, nio* [ni:ə]. 10-20: *tio* [ti:ə], *elva, tolv, tretton, fjorton, femton, sexton, sjutton, arton, nitton, tjugo*. 100: (*ett*)*hundra*. 1000: (*ett*)*tusen*. The number 0 is invariably rendered as *noll*.

² *Femtonde århundradet/seklet* may be used but it is at least as unusual as 'the fourteen hundreds' is in English and it strikes the native listener as abstruse – you have to think for a while to catch the meaning.

struction is practical as an element in compounds, e.g. *1200-talsled* 'channel or path from the 13th century', *1700-talsarkitektur* 'architecture from the 18th century'.

The approaching shift of millennium has actualized the question how higher dates and periods of dates should be pronounced. According to the Swedish Language Committee (Svenska språknämnden) one should not use *tvåtusentalet* but instead *tjugohundratalet* as name for the next century (I designate in what follows the two ways of pronunciation as 2.000 and 20.00, respectively). Many Swedes oppose strongly the recommendation of the committee and I will also raise some objections to the official notion. In this connection, I will discuss the construction of numerals. Some other Indo-European languages are then studied as a confrontation with my solution(s) to the Swedish problem.

Sigurd 1972:4 talks about certain numerals as representing non-total fusion. Unlike the cases just mentioned, they exhibit division into parts or no division at all – a number is thus treated as a succession of less complex numbers. A section on English numerals contains some cases of non-total fusion.

2 THE SWEDISH CASE

2.1 The 20.00 model

Allén 1981 argues for the reading 20.00 – which thus gets a certain official mark through Allén's membership in the Swedish Language Committee and the Swedish Academy (*Svenska Akademien*). He starts off with the statement that Swedish lacks an established denomination for the period 1900-9. He further remarks that the other decades have been given names after the last two digits in their respective first year. According to the same principle, the first decade would then get the name *00-talet*. A rightful objection to the comparison is that the denominations are not quite parallel. *Tiotalet* is in Sigurd's terms characterized by fusion, while *nollnolltalet* is not fused at all. The nearest possible correspondence to *nollnolltalet* is not *tiotalet* but *ettnolltalet*. A totally paradigmatic solution to the problem of naming the decade would be *nolltiotalet*, but the form seems even more ungrammatical than Allén's suggestion and would surely have difficulties in getting accepted in colloquial speech. Observe that *nolltio* ought to be pronounced as [nɔlti] to match the upper tens (which normally end with [ti]).

Allén presents some arguments and quotations for the designation *nollnolltalet*. A good argument is of course that the year 1900 apparently may be called 00, especially in connection with persons born in that year. In the paper he further mentions that Tigerstedt, in his history of literature (1971), says that the word *tioalister* is an improper term for a number of writers and is only motivated for practical purposes. Behind this lies apparently the difficulty in

finding a useful designation for the period 1900-9. Allén shows convincingly that Tigerstedt not just accepts that the generation of writers from that period for language practical reasons are named *tioalister*, but that he also to some extent (unconsciously, to be sure) displaces the imagined essence of their writings to the actual *tiotalet* [the second decade].

Ellegård 1983 hardly believes that the name of the first decade is acceptable in the (unfused) form *nollnolltalet* or even (with a new variant) as *nolltalet*. In his opinion, *tiotalet* designates by interference not just 1910-19 but 1900-9 as well.³

I think the last statement is a misinterpretation. Allén's example rather shows that Tigerstedt avoided trying to denominate the first decade. The term *tiotalet* presumably aims generally at the period 1910-19. The probable reason for avoiding *nollnolltalet* will be mentioned later in this paper.

Allén also refers to Strindberg, who according to a source did use the form *nollnolltalet* at one time. Yet Strindberg was not only the writer of important *œuvres* such as *Ett drömspel* and *I havsbandet*. He entertained far-reaching interests beyond authorship and not everything he touched turned into gold. Interestingly, his linguistic writings (1920) keep the same low standard as the works of many other etymology amateurs.

Strindberg was an expansive story-teller, who enjoyed stretching words and phrases to see how far they lead. Therefore, the half intimated argument that a Strindbergian form unconditionally must have a sanction in the future does not work.

Allén notices that the alpha-numeric writing (*x-talet*, where x is at least divisible by 100) leads to problems for his division. The term *900-talet* then becomes ambiguous, as it aims at both a century and its first decade. He proposes, without any motivation, that an apostrophe after the hundred digit should be used to mark the decade.

Then Allén follows the chain of centuries from our time and downwards and suggests that *1000-talet* as the denomination of a century ought to be called *tiohundratalet*, thus conforming to the paradigm he is establishing. *1000-talet* as the name of a millennium is then pronounced differently – as *tusentalet*. Further down among the centuries he finds the years 1-99, which then should be pronounced *nollhundratalet*. He moves on to the twenty-first century, concluding that it ought to be called *tjugohundratalet*. In sum, then, Allén wants to name every first decade etc. by integrating the zero in the system. He differentiates between millenniums and centuries, as he makes use of the formula *x-hundratalet*

³ This interpretation reminds one of Jost Trier's 1973 paradigmatic field theory. Trier vindicated that a word takes on its meaning through its relations to the other words in a semantic field. If a word disappears (in our case: is missing) no lasting semantic hole will emerge. Instead the change will lead to a change in the semantic field, which makes the hole to be filled.

for all centuries. Thus e.g. *tvåtusentalet* means 'the third millennium' in his system.

As for proper dates, Allén 1981:15 says only that 1000 and 2000 should be pronounced with *hundra* as base in the unmarked case but that *tusen* may be used when "you adopt the perspective of the millenniums".

Åke Jonsson, language planner at the Swedish national radio, gives in a radio programme (1988) some information on how he and his colleagues consider that dates should be named. He mentions that the Language Planning Group (*Språkvårdsnämnden* – with representatives from e.g. the Swedish Academy and the Swedish Language Committee) in a report 1984 proposed that the year 2000 might be pronounced differently depending on the speaker's aim "but after that one should thus count 20.01, 20.02, 20.03 and so on". Regarding the proper dates Allén's fine distinction between hundreds and thousands, which Jonsson proudly explains that some radio and TV people and lately even other official persons have "learnt", is apparently not valid. Jonsson does not state the reason why e.g. 2.001 is not recommended besides 20.01, but a likely motive would be the relative rarity of the millennium thinking in such cases. Another explanation is that favourable conditions for the term 20.00 might cease to persist if variation were allowed in the pronunciation. Jonsson bears out the latter explanation when he assumes

that the option in the first case does not really hold, if you want to see the use of 20.00 carried out. If we have got used to saying 2.000 throughout the first starting year it will probably demand too much of us if we then should pass over to 20.01 etc. The best must be to (during the New Year festivities in the year 1999) prepare oneself for the fact that – in a few hours – a new century begins, and the first year in that century is called 20.00.

Apart from the fact that Jonsson vindicates a somewhat harder line than the Language Planning Group, he agrees with them that the other dates in the next century ought to be called 20.01 etc. The parallelism between hundreds and thousands is thus not upheld at the level of dates and it may be assumed that the thinking in thousands is irrelevant here – and therefore should be so with period designations as well. Historians do not seem to look upon the millenniums as especially suitable units either, but prefer to use the pronunciation 1.066 when speaking of the battle at Hastings. We should therefore question the opinion on behalf of the Language Planning Group that (ibid.)

it might be practical to differentiate between thousands and hundreds when we in a little more than ten years face a shift in both these cases.

Because there does not seem to exist any need for special short denominations for the millenniums, that argument would fall and we may set 2.000 free. In 2.3, I motivate why this form should be preferred to 20.00.

When the Danish language committee in 1968 was asked how the next century should be pronounced, it passed the question to the Norwegian and Swedish language committees to let them make additional statements. The Danish and Norwegian committees opt in their answers for the solution 2.000 but do not oppose the pronunciation 20.00. The Swedish Language Committee – which responded first – recommends 20.00 without reservations. The reasons they convey are "partly that *1000-talet* **alternatively** is read as *tiohundraalet*, partly that *tvåtusentalet* may designate the entire third millennium [my emphasis]" (Allén 1981:16). The paragraph issues for some reason in the rhetorical question whether the whole of Scandinavia ought not agree on the Swedish proposal.

Ellegård 1983 mainly concurs in the Allén analysis, which he does not believe that "anyone may effectively challenge", but asks himself why it anyway feels more natural to say 2.000. His answer is that we have not come far enough in time to be completely sure on the systemacy of hundreds and thousands.

For certain reasons, which I refrain from giving here, I do not believe in Ellegård's hypothesis but he is of course right when he writes that cardinal counting with periods of dates is the origin of the discussions about 2000. The power of the non-maximal forms is likely to be reinforced by their presence in the names of the centuries. No discussions have arisen in languages where centuries normally are classified with the aid of ordinals. According to Oscar Lazar only *zweitausend* etc. appear in German radio.

2.2 Earlier criticism of the 20.00 model

So Allén proposes the name 20.00 for the next century and the Swedish Language Committee agrees with him in this matter. The cooperators at the Swedish national radio are told to pronounce the dates in the officially prescribed way and guests in the studios are expected to follow the regulations.

But the resistance against the reorganization seems to be deeply rooted and rather widespread. Laymen have presented some arguments and empirical observations which indicate that the official pronunciation is counter-intuitive and ungrammatical. The debate has even lead to a question in the parliament (Alfredsson 1984).

Arne Ericsson reports in a radio chronicle (1987a) an interesting slip of the tongue from an atomic energy antagonist, who demanded that the nuclear station at Barsebäck should be closed down

in any case absolutely at the latest in the year *tjugotusen* [twenty thousand], we do not want to wait until the year *tjugotusentio* [twenty thousand ten].

I have heard about a similar slip of the tongue, when a female journalist talked about the year *tvåhundrausen* [two hundred thousand], meaning 2000.

These examples show that the combination of *tjugo* and *hundra* leads to problems for the speakers. In the first case the speaker has already produced the first element and solves his inner linguistic conflict by interchanging *tusen* for *hundra*. The second case may be regarded as another instance of raising powers of ten, but here it takes place at an earlier and more abstract stage in the generation – where the representation is $/2 \cdot 10^1 \cdot 10^2/$. After the raising of the first power of ten, the other must be raised as well, because languages do not generally tolerate sequences of phonologically identical morphs (cf. Menn and MacWhinney 1984; Hurford 1987:260f.). This is thus clear evidence that the speakers hesitate when confronted with the officially prescribed dates and I think the explanation I have provided is tenable.

Ericsson *ibid.* mentions that the year 10000 would be pronounced *hundrahundra* according to the official view, a pronunciation that he finds

illsounding and hardly understandable. The problem lies certainly to a high degree in the future – but one should as has been said consider the question carefully (that is the opinion also of language planners within the national radio).

It is of course a great disadvantage that the official model might produce a form like this, that would be overridden by the general constraint against repeated morphs. A form like *hundra tvåhundralet* also sounds odd, as it can be interpreted as either 'the 103rd century' or as an approximate mentioning of time – 'the second and/or third century'. Some other constructions, like 103.00, merely sound strange in the latter interpretation, but seem nonetheless ambiguous. The restriction on identical morphs in a row seems then to be more severe than what was stated earlier, in the case of the numerals. In a numeral syntagm XABAY, where the two A morphs are phonologically identical and where X and Y may contain material, B is bound to contain a power of ten which is bigger than A. This suggested restriction should of course be checked more carefully.

Note that as the paradigm *x-hundra* theoretically takes on values for *x* from 0 and upwards it manages to produce several ungrammatical forms. In the next section I will try to show that also e.g. 20.00 is ungrammatical.

2.3 A rule for generating dates and an explanation for the origin of the special cases

I remarked in the introduction that the Swedish dates to a certain extent violate the packing strategy or rather work according to a specific rule. To find this rule

we also must consider the main rules for the construction of numerals. Sigurd 1972 presents a rather satisfactory description, which will be reproduced in parts. The major change that has been done in the transmission is that the mathematical background has been emphasized in that the words *tio*, *hundra* etc. are written in the form of powers of ten. The rule system is thought of as a production model and does not reflect the understanding of numerals. The present object is anyway the building of the complex numerals.

The atomic morphemes which mathematically correspond to positive powers of ten will henceforth for the sake of simplicity be called numeral measures. The denomination is based on the fact that these morphemes are quantity units that may take other numerals as attributes. The self-referring feature of the numeral measures may seem a bit strange but the numeral group has on the whole quite special properties.⁴

Sigurd first gives some deep structure rules:

- 1 a) $N \rightarrow M \cdot 10^6 + M \cdot 10^3 + M$
- b) $M \rightarrow d \cdot 10^2 + d \cdot 10^1 + d$
- c) $d \rightarrow \{1, 2, 3, \dots, 9\}$

N represents any complete numeral and M a mediator (a construction which may at the same time be attribute to a numeral measure and contain one or more numeral measures). The set representing the natural numbers that are smaller than the smallest numeral measure (and which in written form are marked with digits) is called d. The existence of a numeral measure appears in writing indirectly from the position system, where the zeros fill the function of placeholders.

The rules may be written more simply and (in the first case) more adequate with small changes. We obtain the following rules:

⁴ The denomination has also support from historical linguistics. Menninger 1958:136-44 derives the old Indo-European word for 10 – *dekṃ* – from a combination of the elements with the meanings 'two' and 'hand' (cf. *ten*, *two* and *hand*) and shows in addition that the nearest higher numeral measures trace their origin back to this old Indo-European compound. Swadesh 1971 gives another explanation for *dekṃ*, but notes as well (p. 84) that 10 in many languages has the (at least etymologic) meaning 'two hands'. The decimal position system is therefore not an arbitrary system which happened to strike through by accident, but its dominance has a prerequisite in finger counting. The higher numeral measures were originally constructed on a basic unit but have later been truncated. We may compare this with units of weights, where in Swedish two of them are called *hekto* and *kilo* as the last element – *gram* – has dropped out. Bengtson 1987 derives IE '10' from a word meaning 'hand' and IE '100' from another word, though with a similar meaning – 'arm' or 'hand'.

- 2 a) $N \rightarrow (M+10^3 (M+10^3 (M+10^3 \dots)))$
 b) $M \rightarrow (d+10^1 (d+10^1 (d)))$
 c) $d \rightarrow \{1, 2, 3, \dots 9\}$

Lexical rules specify e.g. 10^9 as *miljard*. The pronunciation of the higher numeral measures is dependent upon the specialization in the vocabulary of each individual speaker.

The number 0 does not enter into the set *d*, something which contradicts Allén's assertion that e.g. *nollhundraalet* exists as a possibility. 0 holds anyway a unique position among the natural numbers, to which it at times is assigned. It neither belongs to the positive nor to the negative numbers; division with 0 is forbidden (the quotient is not defined); if 0 participates as a term in multiplication with two or more terms the product becomes 0.

Linguistically, 0 also has special qualities. We do not talk about the year 0, as Allén himself (p. 15) writes. Already Cederschiöld 1897:24 explained that the use of *noll* as proper cardinal is very limited, especially in the adjectival and attributive functions. In Olsson 1988 I show that *noll* is odd as part of numeral constructions and makes them unanalyzable.

Similar arguments against the existence of 0 in the set *d* are delivered by Hurford 1987:95. He rejects the set theoretical axiom "Zero is a natural number" with the motivation that numeral systems in natural languages do not contain 0. The number which the untrained speaker intuitively states as the smallest number is 1. There is abundant evidence that the mathematician Peano – to whom Hurford ascribes the above axiom – also rightly gave 1 as the first natural number – cf. Peano 1908:37, Johnstone 1987:28. He adds that 0 in most languages is not an element in compounded numerals. Sentences with *zero* or *nought* as attributes would be ungrammatical in English – Englishmen do not say e.g. *We have nought bananas* or *We have zero bananas* but on the other hand *We have no bananas*, where *no* however is not regarded as a numeral.⁵ Then there follow

⁵ An Englishman told me, however, that *zero* is normal as attribute and predicate in American English – at least in some dialects of it. An American English informant gave the following example, that she once experienced, on this matter: *She has zero patience*. She added that such sentences are extremely common, but there is a restriction on the occurrence of *zero* as attribute: it must be followed by a really uncountable head, i.e. an abstract noun. A noun like *water* is not abstract: it is within normal human ability to measure a given amount of a liquid, in quite another way than the degree of e.g. *patience* – while numerals can not stand bare in front of any of these items, they may be attributes to attributes of nouns like *water* (as in *two gallons of water*). The noun *water* appears, furthermore, in a plural form – unlike *patience*. An American English speaker may use *no* in front of both items, but before *water* this is the only option. To conclude, *zero* and the numerals are in complementary distribution – at least in attributive

generation rules which are closer to the linguistic surface. Some of the so-called lexical rules are:

- 3 a) $1(10^1) + 1 \rightarrow \text{elva}$
 b) $1(10^1) + 2 \rightarrow \text{tolv}$
 c) $1(10^1) + d \rightarrow \text{d+ton}$
 d) $2(10^1) \rightarrow \text{tjugo}$
 e) $10^1 \rightarrow \text{tio}$

The rules are in principle identical with Sigurd's 1972:18 except that they have been more formalized. I order the rules a and b before c as they fulfill its structural description but are special cases which do not obey the main rule. In accordance with the same system, the very general rule e comes last among the rules in 3.

The next step is to construct a rule that generates the special date forms.

What happens in the Swedish cases is that *hundra* replaces *tusen* as highest numeral measure. The unit that is an attribute to *hundra* is combined with a ten (I will from now on call such syntagms ten expressions). The formalization is simple:

$$4 \quad d_1(10^3) + d_2(10^2) \rightarrow (d_1(10^1) + d_2)10^2$$

It looks almost as if the higher numeral measures were rearranged so that the entire value should be expressed in one of them, which may also be part of the reason for the non-maximal forms (there is no reason here, either, for a change from 2.000 to 20.00).

The rule fits in between the deep structure rules and the lexical rules. Later on the *ton* conversion rule 3c is at work.

The automatic generation does not create any form like 20.00, because 0 (with which the second power of ten would have to be multiplied) does not belong to the set *d*. It is interesting to note that the rule generally allows conversion on the

position. The abstract nouns may take *zero* as attribute, but not a proper numeral. In terms of Coseriu's 1967 lexical solidarities, the abstract nouns constitute a class which determines through affinity the possibility for *zero* to act as an attribute.

only assumption that both powers of ten are realized (i.e. that a *d* precedes each one of them). Higher ten expressions may thus be attributed to *hundra*. Sigurd 1987:119 (cf. also 1972:7) says that teens are special and assumes that they – unlike higher numerals – are allowed to violate the packing strategy. It is naturally true that teens are special but I have at least shown that theoretical reasons speak in favour of higher ten expressions being acceptable attributes to *hundra* because such syntagms are generated with the simplest rule. Hurford 1976:72 says that *twenty eight hundred* and other instances of *hundred* with a higher ten expression are perfectly acceptable in English.

Generative grammar thus explains the technical details of the problem, but the deeper meaning – the historic (and presumably also synchronic) causal connection – is found in a natural phonological tendency. The explanation is not difficult to catch. Swedish, like some other languages, here makes use of the possibility to shorten unnecessarily long dates. Allén 1981:15 grazes the explanation, but does not comment on it. He just says that 1984

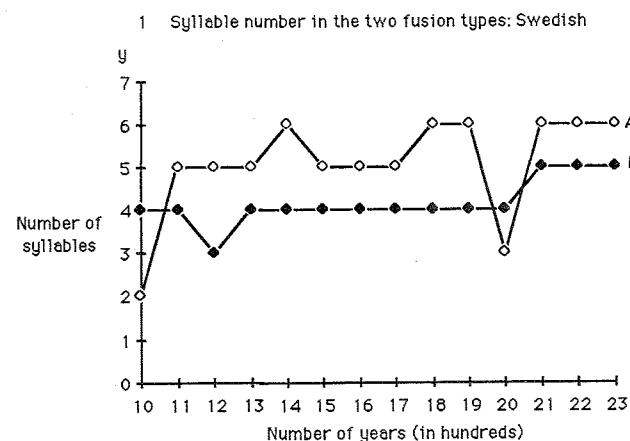
as date [...] is called, three syllables shorter, *nittonhundraåttiofyra*. We may thus observe one numeral principle and one date principle.

The following diagram shows where the non-maximal fusion will result in a shortening (naturally, only the first year of a century is needed in the description, as the other dates in the century exhibit the same pattern, their additional parts being invariant). The syllable has been taken as time unit. Maximal fusion is illustrated by line A and the non-maximal fusion where *hundra* is the highest numeral measure is symbolized by line B. I presume that *ett* normally precedes *hundra* but not *tusen* in these cases. The analysis of *tjugo* is built upon a non-reduced pronunciation with two syllables, not one (the calculation fits in on the other higher ten expressions but the difference between A and B might thus be greater in this case).

The diagram shows that if *x* is a ten, then A is made up of fewer syllables than B, while the opposite is true for ten expressions. The reason for the origin of the non-maximal forms should thus be their shorter pronunciation. Allén tries to extend the group of non-maximal forms, but he then strives towards another goal than the original one.

The correlation between the natural shortening principle and the formal conversion rule 4 is clear and points towards the generalization that conversion takes place only if ten expressions then form attributes to *hundra*.

The question about which of the components is most prominent in the single case – the conversion rule or the shortening principle – is hard to decide. In section 4 some evidence will be given in favour of the rule.



3 HURFORD'S SYNTACTIC SOLUTION AND EVIDENCE AGAINST IT FROM OTHER LANGUAGES

Hurford 1975:72 ff. gives another explanation for the non-maximal forms in e.g. Swedish. He thinks that alternants with ten expressions which are followed by *hundred* are preferred to the maximal forms, because their grammatical construction is simpler in that they have a simpler tree structure and thereby are easier to process:

By none of these measures is *two thousand two hundred* more economic than *twenty two hundred*, and by all the node-counting measures the latter is definitely more economic than the former (Hurford 1975:77f.)

He thus explains the problem why the exceptions to the packing strategy appear in terms of their internal structure, which is more economic than in the logically more regular alternatives. He thinks that the measure of economy which gives verdict is defined in terms of the sum of the nodes in the structure, or possibly as the number of nodes in some of the four specific node classes. Hurford *ibid.*:78 adds, though, that the

possibility cannot be ruled out that the correct measure of economy will take into account aspects of surface structure, such as, perhaps, number of stressed syllables. Also by this measure, *twenty two hundred* is more economic than *two thousand two hundred*.

An investigation of the tree structures for 2.000 respective 20.00 in Hurford's terms shows that his hypothesis is correct for English. 20.00 is the costlier alternative in regard to nodes (one more) and has one more 'lexical item'. 2.000

- 2 a) $N \rightarrow (M+10^3 (M+10^3 (M+10^3 \dots))$
 b) $M \rightarrow (d+10^1 (d+10^1 (d)))$
 c) $d \rightarrow \{1, 2, 3, \dots 9\}$

Lexical rules specify e.g. 10^9 as *miljard*. The pronunciation of the higher numeral measures is dependent upon the specialization in the vocabulary of each individual speaker.

The number 0 does not enter into the set *d*, something which contradicts Allén's assertion that e.g. *nollhundralet* exists as a possibility. 0 holds anyway a unique position among the natural numbers, to which it at times is assigned. It neither belongs to the positive nor to the negative numbers; division with 0 is forbidden (the quotient is not defined); if 0 participates as a term in multiplication with two or more terms the product becomes 0.

Linguistically, 0 also has special qualities. We do not talk about the year 0, as Allén himself (p. 15) writes. Already Cederschiöld 1897:24 explained that the use of *noll* as proper cardinal is very limited, especially in the adjectival and attributive functions. In Olsson 1988 I show that *noll* is odd as part of numeral constructions and makes them unanalyzable.

Similar arguments against the existence of 0 in the set *d* are delivered by Hurford 1987:95. He rejects the set theoretical axiom "Zero is a natural number" with the motivation that numeral systems in natural languages do not contain 0. The number which the untrained speaker intuitively states as the smallest number is 1. There is abundant evidence that the mathematician Peano – to whom Hurford ascribes the above axiom – also rightly gave 1 as the first natural number – cf. Peano 1908:37, Johnstone 1987:28. He adds that 0 in most languages is not an element in compounded numerals. Sentences with *zero* or *nought* as attributes would be ungrammatical in English – Englishmen do not say e.g. *We have nought bananas* or *We have zero bananas* but on the other hand *We have no bananas*, where *no* however is not regarded as a numeral.⁵ Then there follow

⁵ An Englishman told me, however, that *zero* is normal as attribute and predicate in American English – at least in some dialects of it. An American English informant gave the following example, that she once experienced, on this matter: *She has zero patience*. She added that such sentences are extremely common, but there is a restriction on the occurrence of *zero* as attribute: it must be followed by a really uncountable head, i.e. an abstract noun. A noun like *water* is not abstract: it is within normal human ability to measure a given amount of a liquid, in quite another way than the degree of e.g. *patience* – while numerals can not stand bare in front of any of these items, they may be attributes to attributes of nouns like *water* (as in *two gallons of water*). The noun *water* appears, furthermore, in a plural form – unlike *patience*. An American English speaker may use *no* in front of both items, but before *water* this is the only option. To conclude, *zero* and the numerals are in complementary distribution – at least in attributive

generation rules which are closer to the linguistic surface. Some of the so-called lexical rules are:

- 3 a) $1(10^1) + 1 \rightarrow \text{elva}$
 b) $1(10^1) + 2 \rightarrow \text{tolv}$
 c) $1(10^1) + d \rightarrow d+\text{ton}$
 d) $2(10^1) \rightarrow \text{tjugo}$
 e) $10^1 \rightarrow \text{tio}$

The rules are in principle identical with Sigurd's 1972:18 except that they have been more formalized. I order the rules a and b before c as they fulfill its structural description but are special cases which do not obey the main rule. In accordance with the same system, the very general rule e comes last among the rules in 3.

The next step is to construct a rule that generates the special date forms.

What happens in the Swedish cases is that *hundra* replaces *tusen* as highest numeral measure. The unit that is an attribute to *hundra* is combined with a ten (I will from now on call such syntagms ten expressions). The formalization is simple:

$$4 \quad d_1(10^3) + d_2(10^2) \rightarrow (d_1(10^1) + d_2)10^2$$

It looks almost as if the higher numeral measures were rearranged so that the entire value should be expressed in one of them, which may also be part of the reason for the non-maximal forms (there is no reason here, either, for a change from 2.000 to 20.00).

The rule fits in between the deep structure rules and the lexical rules. Later on the *ton* conversion rule 3c is at work.

The automatic generation does not create any form like 20.00, because 0 (with which the second power of ten would have to be multiplied) does not belong to the set *d*. It is interesting to note that the rule generally allows conversion on the

position. The abstract nouns may take *zero* as attribute, but not a proper numeral. In terms of Coseriu's 1967 lexical solidarities, the abstract nouns constitute a class which determines through affinity the possibility for *zero* to act as an attribute.

has on the other hand a larger maximum depth and therefore the two expressions are equivalent so far in respect to economy. Because of the packing strategy the speakers then choose 2.000.

Some data from other languages indicate that the tree structure solution (both universally and from a historical point of view) explains the connection less satisfyingly than the phonological solution.

In Swedish the date 1600 is always rendered as *sextonhundra*. A morpheme-for-morpheme-translation of the Swedish word into Finnish would be pronounced as *kuusitoistasataa*, but no Finn makes use of this form. Instead Finns use the form *tuhatkuusisataa*, which would correspond to a Swedish *tusensexhundra* (*tuhat* means 1000 and *sataa* 100). According to Hurford's hypothesis, the Swedish form appeared because the tree structure was considerably less costly than in the form 1.600. But there is a direct parallel here in Finnish, which yet makes use of maximal constructions in all current forms. Finnish even has a special word, that resembles the Swedish *ton* in that it does not mean 10 except in additive connection with a unit (*toista* is by the way partitive singular of *toinen* 'other'). The difference between the Finnish and the Swedish ways of treating dates gets a very simple explanation if we instead handle the problem phonologically, by applying the shortening principle. The forms *tuhatkuusisataa* and *kuusitoistasataa* both contain six syllables. Because the non-maximal forms do not follow the packing strategy and do not lead to any economic advantage either, there is no motivation for loading the grammar with a rule that generates them.

Yasuko Nagano has informed me that Japanese neither allows the non-maximal forms. 16.00 – *juuroppyaku* – is an ungrammatical form which is by no means current (unlike *ton*, *juu* corresponds to 10 in all positions). The correct structure is here 1.600 *senroppyaku*, because the pronunciation does not matter as both expressions are pentasyllables.

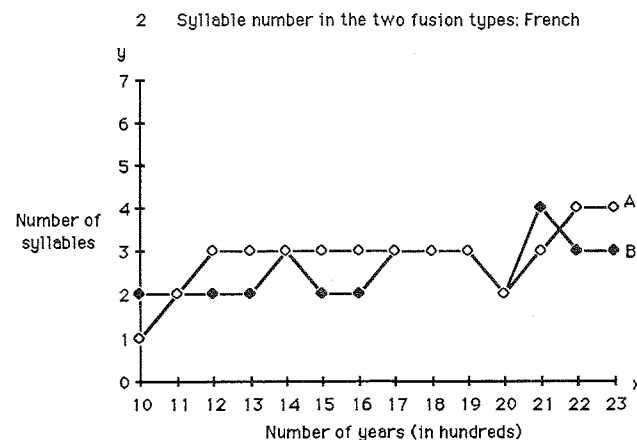
The explanation to the fact that the to us well-known non-maximal forms are not used in these two languages should therefore be that the forms do not convey any relaxing in the pronunciation.

4 NON-MAXIMAL NUMERAL TYPES IN OTHER LANGUAGES

4.1 The optional use of non-maximal numerals

It seems that there are languages which have rule 4, but use it optionally. One such language is French. I devised an inquiry to find out how native speakers of English pronounced dates (see the next subsection). This inquiry was slightly changed and presented to three Frenchmen, who differed considerably from the English informants in their answers (a difference which apparently is more a case for cultural research was the fact that whereas the English speakers filled in

the inquiry, the French informants preferred to pronounce the forms). According to the informants, dates like 1400 may be pronounced as *mille quatre cents* or *quatorze cents*, 1700 either as *dix-sept cents* or as *mille sept cents*. To be more specific, French makes optional use of rule 4. The syllable length of French numerals in the two fusion types is given on the following chart.

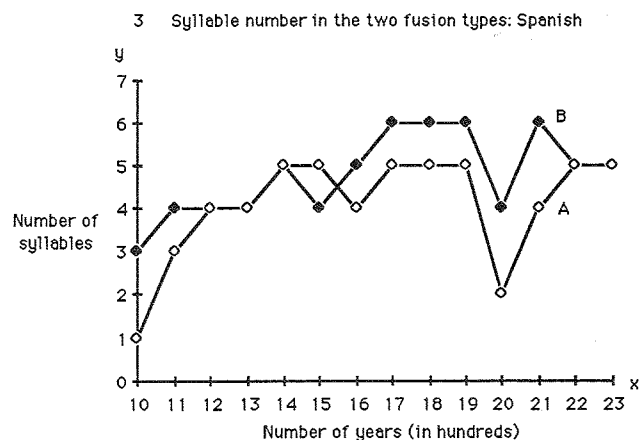


Note that the difference in syllable length between the two types is smaller than in Swedish, which might explain the variation. I then looked at another Romance language, namely Spanish. Counting by hundreds is, according to Harmer and Norton 1946:44, "used in Spanish up to 900 [sic!] only: *mil novecientos treinta y tres*, 1933". Bejarano and Jörnving 1973:61 note the same fact:

Observe the difference between Swedish and Spanish with dates: *mil novecientos sesenta y seis* nittonhundrasextiosex.

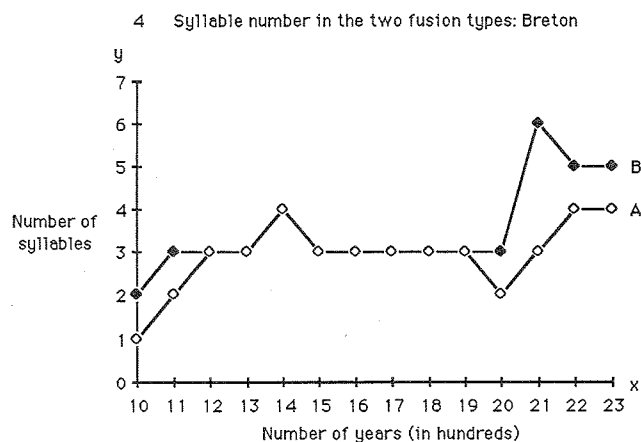
The hypothesis predicts, then, that in Spanish there will be no economic advantage of using the non-maximal numerals. This is borne out, as can be seen on the following chart.⁶

⁶ Spanish 1-9: *uno, dos, tres, cuatro, cinco, seis, siete, ocho, nueve*. 10-15: *diez, once, doce, trece, catorce, quince*. Higher ten expressions are constructed with the formula $d \cdot 10^1 + d \rightarrow d \cdot 10^1 y d$ (which might be simplified), though the spelling now has *dieciséis* etc. for teens and twenties (20 has here the form *veint* and an accent mark is placed on the first vowel of a closed final syllable). 20: *veinte*. 100-900: *ciento, doscientos, trescientos, cuatrocientos, quinientos, seiscientos, setecientos, ochocientos, novecientos*. The presented vowel combinations all consist of one syllable.



In Spanish, maximal numerals are on the whole shorter than their non-maximal counterparts or of equal length.

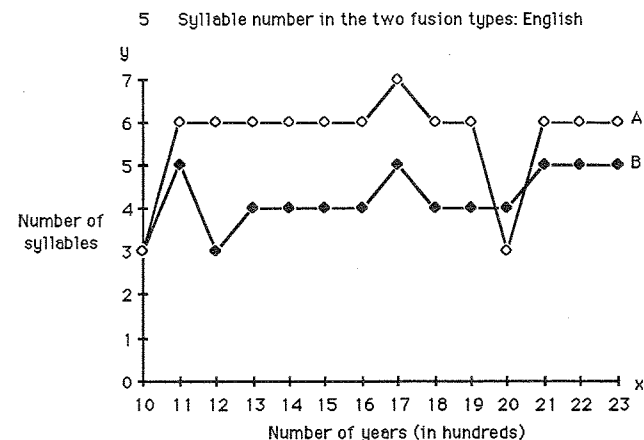
So far so good. The hypothesis has been supported, also because the internal structures of these Romance languages are very similar, so that any syntactic or morphological solution seems to be out. But proceeding to Breton the picture emerges at once less clear. Breton is reported as having alternations of the French type (Press 1986:88), which should mean that it is more economic to use non-maximal numbers here. But as the following chart shows, Breton is like Japanese and Finnish in this respect – and yet rule 4 is at work, even though it is optional.



I can not provide an explanation for these facts, unless the French system has been transferred into Breton.⁷

4.2 Non-maximal numeral types in English

English numerals in the two fusion types are shown on the following chart.



The same picture emerges as in Swedish, so one would expect rule 4 to be active, but there are some additional facts in English which contradict this assumption. As mentioned in the preceding section, I made a small inquiry to see how five English speakers would pronounce different numbers with three or four digits.⁸ The answers to this inquiry indicate that fusion is less prominent in English than in Swedish. When a number is of the type $dx0d$, where x may be either d or 0 , the first two digits are fused and 0 is rendered as $o(h)$ [əu] (this fact would conceivably originate in a spelling pronunciation). A number of the type $dxdx$ exhibits only internal fusion in the dx -constituents, without any link between the pairs. A general rule inserts *hundred* in $dd00$, while *thousand* is generally found only in $d000$. Skipping, for a while, the question as to how the rule converting

⁷ Breton 1-9: *unan, daou, tri, pevar, pemp, c'hwec'h, seizh, eizh, nav*. 10-19: *dek, unnek, daouzek, trizek, pevarzek, pemzek, c'hwezek, seitek, triwek'h, naontek*. 20: *ugent*, 20 + d is d *warn-ugent*, where *warn* means 'on the'. 100 is *kant* and 1000 is *mil*. 200 is *daou-c'hant*; *tri, pevar* and *nav* also take *c'hant* instead of *kant*. The presented vowel combinations all seem to consist of one syllable.

⁸ The instructions read as follows: "Would you please give these dates in the written form that, according to your judgement, corresponds best to their actual pronunciation?" The dates were 1373, 1700, 1508, 1008, 1090, 1066, 2053, 2004, 2070, 2304, 2000, 2362, 605 and 923. In the inquiry to the French speakers, 1462 was used instead of 1373 and 1984 and 1632 were added at different places.

10^3 to 10^2 should be written, we proceed to the more intriguing facts about the non-total fusion in English dates.

Assuming that \emptyset and o are variants of *hundred*, the following ordered rules will generate all three of them correctly.

- 5a) $10^2 \rightarrow \emptyset / _ d \cdot 10^1$
 b) $10^2 \rightarrow o / _ d$
 c) $10^2 \rightarrow \text{hundred}$

Rule a applies to a subset of the input to b, which it therefore bleeds. Rule b bleeds similarly c.

It is obvious that rule 4 does not cover all the English cases, because there are forms like *twenty seventy*. It is possible to relax rule 4 so as to permit a transition to 10^2 as highest power of ten, if only some position after *thousand* is filled with a d (the attribute to a hundred is still combined with the attributive ten):

- 6 $d_1(10^3) + d_2(10^2) \rightarrow (d_1(10^1) + d_2)10^2 / _ d_3$
 where d_2 or d_3 , but not both, may be absent

The rule may be simplified to

- 7 $d(10^3) \rightarrow d(10^1(10^2)) / _ d$

Hundred will occur only if the rightmost d is (part of) the attribute of 10^2 and then, in order not to violate the principle at the end of subsection 2.2, the two instances of 10^2 will be joined (a simplification to the same effect may of course be performed on rule 4).

This is the colloquial way of constructing English numerals and it is not clear to me whether it has somewhere been described instead of the standardized but in practice rather unemployed system (showing total fusion) with which e.g. Hurford 1975 is concerned (one informant said that longer forms are strongly marked in most cases, but thinkable – especially within metrical lyrics).

Hurford 1987, Chapter 6, remarks that numerals are subject to standardization, but that there may be variation. This seems especially true in the case of English, where the inquiry showed considerable diversity. For the forms 1008 and 2004 the pronunciations *one thousand and eight, ten hundred and eight* (without rule 5b) and *two thousand and four* were also found (*and*-insertion

seems to be optional in some dialects, as Hurford 1975:51 states). These forms were only used by the women in the investigation, and not even then exclusively. Two informants had *six hundred and five*. One of them had also *nine hundred and twenty three*, alongside *nine two three*.

The use of hyphens is probably country (i.e. education) specific: the British speakers did not use hyphens, whereas the North American speakers used it everywhere or according to special rules.

As a final point, I wish to emphasize that $o(h)$ in a way behaves like and must be treated as a real numeral. It is part of numeral constructions, denoting 10^2 , while at the same time it is clearly connected with the figure 0. The spelling is surely in a sense responsible for the pronunciation, since o is also one of the various realizations of the number 0. There seems to be a prohibition against it being repeated, cf. *double o seven* for 007. The corresponding Swedish form is *nollnollsju*. According to an informant, o is not repeated but the construction *double o one* etc. is used in telephone numbers and the like (a sequence consisting of four zeros is rendered as *double o double o*, but it is not clear if an odd numbered sequence of zeros has to be pronounced without repetition – e.g. if a sequence of three zeros is rendered as *o double o* in all cases or even at all). The same goes for repeated instances of other digits. Swedish does not use the same way of expressing sequences of zeros; a postal code may be rendered as *tvåhundrafyrtiofyranollnoll* (244 00), while in a six digit telephone number the sequence 99 is fused as *nittionio*. The English informant said that the formal time expression 24.00 is pronounced *twenty four hundred (hours)*, while in Swedish the zero part is simply pronounced *nollnoll* (and *timmar*, meaning *hours*, is not used).⁹ Thus the prohibition seems to exist in English, presumably because o is felt to be a numeral.

5 CONCLUSION

It seems that in the case of the years 1100-1999 the non-maximal fusion became prevalent because it resulted in a smaller number of syllables (the assumption that a seemingly syntactic problem should have a phonological solution might seem shocking to some). Allén wants to incorporate 2000 in the non-maximal group for other reasons – 2.000 is by all means the shorter form. His opinion is based upon a paradigmatic vision, where all assumed inequalities in the pattern have been erased. The present analysis explains why Swedes yet seem to prefer the 2.000-model, if they have the possibility to choose. The pattern is also, as I have shown, more regular in the suggested model than in Allén's model. A strong support for my hypothesis is – finally – the fact that the conversion rule

⁹ The English solution is however legion in military speech.

can not generate the combination $d \cdot 10^1 \cdot 10^2$ (where a ten precedes *hundra*), because 0 does not belong to the set d .

The French and Spanish data show that the rule is generalized to cover all cases. The English rule 5a would considerably shorten the forms in the other languages as well, but these languages apparently lack it. Thus the principle of shortening is only part of the origin as there must come about one or more morphosyntactic rules to effectuate it.

Only future can tell which system the Swedes will use, but I think the arguments presented here deserve attention. It is possible that the Allén model will eventually dominate (several Swedes have told me they think it is only a matter of time or, alternatively, that the Allén model is already established). The fact that it seems to be unnatural does not contradict this conclusively. Hurford 1987:84f. reports that the Welsh numeral system has been regularized by some authority or corporation to fit the decimal position system. The teens in this standardized system have the word for 1 before the word for 10, so 19 is pronounced as *un deg nav*, literally 'one ten nine'. In other languages, this combination is not normally used. Even in Burmese numerals, which Sigurd 1972:30 says "look almost as if they were constructed at the writing-table", *ta* 1 is only optional before *hse* 10. *Ta* is obligatorily deleted in the teens, so 19 is pronounced only as *hse.kou*. But there is little reason why the artificial Welsh system should undergo a change and *un* be eliminated before *deg*. However, I imagine that the Swedish case is different for reasons of grammaticality.

Ericsson 1987b writes that

In the name of reason the formulation "är tvåtusen" will probably win in the colloquial usage, whatever radio and TV might say.

I think this is a correct evaluation of the situation even if the Swedish Language Committee tries to force its own norm of pronunciation through, a norm that is likely to be ungrammatical.

REFERENCES

- Alfredsson, Hans O. 1984. 'Fritt att välja uttal på år 2000'. *Svenska Dagbladet* May 26, 7.
- Allén, Sture. 1981. 'Tjugohundratalet och några andra tidsord'. *Språkvård* 1981:2, 13-16.
- Bejarano, Virgilio and Rolf Jörnving. 1973. *Spansk grammatik*. Uppsala: Almqvist & Wiksell.
- Bengtson, John D. 1987. 'Notes on Indo-European "10", "100", and "1000"'. *Diachronica* IV:1/2, 257-62.

- Cederschiöld, Gustaf. 1897. *Om grundtalens lexikaliska behandling*. Stockholm.
- Coseriu, Eugenio. 1967. 'Lexikalische Solidaritäten'. *Poetica* I, 293-303.
- Ellegård, Alvar. 1983. 'Det omedvetna motståndet mot tjugohundratalet'. *Språkvård* 1983:3, 19.
- Ericsson, Arne. 1987a. Synpunkten. Med mera – enclosure to *Sydsvenska Dagbladet Snällposten*, April 24-25, 7.
- Ericsson, Arne. 1987b. Synpunkten. Med mera – enclosure to *Sydsvenska Dagbladet Snällposten*, August 14-15, 9.
- Harmer, L. C. and F. J. Norton. 1946. *A manual of modern Spanish*. London: University Tutorial Press.
- Hurford, James R. 1975. *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Hurford, James R. 1987. *Language and number*. Oxford: Basil Blackwell.
- Johnstone, P.T. 1987. *Notes on logic and set theory*. Cambridge: Cambridge University Press.
- Jonsson, Åke. 1988. Ordbruk. The Swedish national radio channel *P1*, April 1.
- Menn, Lise and Brian MacWhinney. 1984. 'The repeated morph constraint'. *Language* 60, 519-41.
- Menninger, Karl. 1958. *Zahlwort und Ziffer. Eine Kulturgeschichte der Zahl*. Göttingen: Vandenhoeck & Ruprecht.
- Olsson, Magnus. 1988. 'Om årtalsbenämningar', to appear in *Nysvenska studier*.
- Peano, Giuseppe. 1908. *Formulario Mathematico*. Torino: Fratres Bocca. Facsimile ed. 1960, Roma: Edizione Cremonese.
- Press, Ian. 1986. *A grammar of modern Breton*. Berlin: Mouton de Gruyter.
- Sigurd, Bengt. 1972. *From mathematical into linguistic representations – and vice versa*. Papers from the Institute of Linguistics, University of Stockholm 15.
- Sigurd, Bengt. 1987. 'Multi language translation of numerals by Prolog'. *Working Papers* 31, 117-35, Department of Linguistics, Lund University.
- Strindberg, August. 1920. *Språkvätskapliga studier*. Samlade skrifter av August Strindberg 52. Stockholm: Albert Bonniers förlag.
- Swadesh, Morris. 1971. *The origin and diversification of language*, ed. Joel Scherzer). Chicago: Aldine & Atherstone.
- Tigerstedt, E. N. 1971. *Svensk litteraturhistoria*. Stockholm: Natur och Kultur.
- Trier, Jost. 1973. *Der deutsche Wortschatz im Sinnbezirk des Verstandes*. Heidelberg: Winter.