THE INTERACTION BETWEEN NUMERALS AND NOUNS

Jim Hurford

1 Introduction

Aims This paper is a descriptive survey of the principal phenomena surrounding cardinal numerals in attribution to nouns, with some concentration on European languages, but within a world-wide perspective. The paper is focussed on describing the syntagmatic distribution and the internal structure of numerals. By contrast, the important topic of the paradigmatic context of numerals, that is how their structure and behavior relates to those of quantifiers, determiners, adjectives, and nouns, does not receive systematic discussion here, although many relevant comments are made in passing. A further necessary limitation in scope is the exclusion of forms which are only marginally cardinal numerals, if at all, such as English *both, dozen, fourscore, pair, triple* and their counterparts in other languages.

Organization. This paper is organized into successive major sections as follows:

1. Simple lexical numerals (Section 2).
2. Simple lexical numerals modifying nouns (Section 3).
3. Complex numerals: internal structure (Section 4).
4. Complex numerals modifying nouns (Section 5).

This organization mirrors a diachronic hypothesis about the emergence of the phenomena in languages. According to this hypothesis, some (but not all) internal structural properties of complex numerals are patterned after the ways in which simple lexical numerals modify nouns. Similarly, aspects of the ways in which complex numerals interact with nouns are dependent on the preexisting structure of the complex numerals themselves and are extensions or adaptations of the ways in which simple numerals interact with nouns. In turn, and looking 'further back', the ways in which simple numerals modify nouns may often reflect structure not involving numerals at all, but attributive adjectives. This is not to say that at each level of development, other aspects of structure may not arise which are sui generis and independent of the prior influences. The facts to be laid out are organized in this way not in a spirit of argument for this broad diachronic hypothesis, but rather in the hope that this organization lends coherence to the whole picture of numeral-noun interaction that will be drawn.

Sample. Given the relatively small number of, and the high degree of historical interrelatedness among European languages, it is probably impossible to arrive at a properly unbiased sample. The languages described here were chosen to give a good geographical spread (North-South, Finnish-Maltese; East-West, Archi-Basque), with no more than three (and usually only one) representatives of any single language group. The languages surveyed here have a Western bias: proportionally fewer languages spoken in the farthest eastern reaches of Europe are discussed. More information is available about some languages than others; the maximum number of European languages mentioned in connection with some pattern for which statistics are gathered is 16, and the minimum number is 10. The coverage is therefore not broad by the standards of modern typological surveys, but interesting phenomena in individual languages or small groups of languages are discussed in some depth. In places, data from non-European languages, especially Hebrew, Sinhala and Shona, are presented by way of contrast with the European data[1].

Several tables are presented, illustrating statistical trends in the data. In all cases, the skewing of the data evident in these tables is, as far as I can see, also roughly representative of the situation across human languages in general. No phenomena unique to European languages have been discovered, except perhaps for:

- Plural group numerals in Finnish (and to some extent Icelandic); in these languages a morphological marking of the numeral X indicates the meaning 'X groups of' (NB not 'groups of X').
- Global morphological marking of complex numerals, which is especially thorough in Russian, Finnish and Greek; in such languages, all, or at least many, of the constituent words in a complex numeral, rather than just the last word, are morphologically marked for a particular feature, such as case, gender or ordinality.

Perhaps the stimulus of such phenomena being described here will provoke their discovery outside Europe. Certainly, other scholars who have worked on numerals have found these phenomena surprising and unfamiliar.

On the other hand, some patterns known from outside Europe may be absent, except in marginal ways, from Europe; for example, no language in this sample has a well developed numeral classifier system.

Within the languages surveyed, it is sometimes possible to see a central/peripheral dimension. Thus, apparently unusual features (for European languages at least)

[1].The languages referred to, with the abbreviations used for them, are: Adyge (Adg), Albanian (Alb), Archi (Arc), Avar (Avr), Basque (Bsq), Bulgarian (Blg), English (Eng), Finnish (Fin), French (Fr), Standard German (Grm), Godoberi (Gdb), Greek (Grk), Hebrew (Heb), Hungarian (Hng), Icelandic (Ice), Kabardian (Kbr), Lezgian (Lzg), Maltese (Mlt), Romany (Rmny), Russian (Rus), Scottish Gaelic (ScGl), Shona (Sho), Sinhala (Sin), Welsh (Wls), Zürich German (ZD).

are often found in languages spoken at the geographical extremities, such as Finnish, Maltese, Basque, Gaelic, Archi and Kabardian. Such unusual features include: lack of a full range of ordinals (Maltese, see section 2.2), plural group numerals (Finnish, see above, and section 3.3.3.3), switch from plural to singular noun after 10 (Maltese and Scottish Gaelic, see section 5.4.1), non-suppletive ordinal for 1 (Archi, see section 2.2), non-adjectival word order specifically for 1 (Kabardian, see end of section 3.2).

Most of the facts reported here were elicited from informants, by the author working with them through a standard questionnaire. Where possible, the information thus gleaned was checked in grammars. In some cases, the information reported comes only from grammars and scholarly articles.

Terminology. It has proved possible to describe most of the data using relatively theory-neutral terms from standard traditional grammar and well-rooted linguistic theory. The most problematic area has been that of agreement, government, and head-modifier relations (see papers in Corbett et al., 1992). In many cases it is clear whether the numeral or the noun is the head, or the modifier, in an attributive numeral-noun construction. But elsewhere it is not so clear, and in order to bring as many languages as possible into comparison with each other, I have simply avoided the issue of whether the numeral, or the noun, agrees with, or governs, the other. What **is** clear, in all cases, is which expression is a noun and which is a numeral, and it has proved possible to make descriptive statements simply in terms of these syntactic categories.

2 Simple lexical numerals

2.1 Arithmetic range of simple lexical numerals

Most languages with counting systems have single morphemes for values up to 10, and thereafter use syntactic combinations to express higher numbers. There are languages with lower bases than 10, usually 5; in such languages, expressions for the numbers 6 - 9 are bimorphemic. I have found no trace of such quinary counting in modern European languages.

In Finnish, one can discern the ('plural') forms for '1' and '2', *yhde* and *kahde* in the forms for '9' *yhdeksän* and '8' *kahdeksan* respectively , indicating a bimorphemic subtractive origin of these numerals, but modern Finns are typically not conscious of the forms for '8' and '9' being bimorphemic, and can asign no meaning to **deksan*. Throughout Europe generally, with few exceptions, the onset of complex syntactic numerals comes at '11'. See section 4.3 for further discussion.

2.2 Distinct counting forms

The numeral which is used to quantify a noun in a noun phrase is not necessarily the same form as the corresponding numeral in the conventional recited counting sequence. I refer here to differences other than those, such as inflections, due to the influence of the sister noun or mother noun phrase. There are various degrees of idiosyncratic difference between a counting numeral and a quantifying (attributive) numeral.

The most extreme difference is where the counting numeral is a **suppletive** variant of the quantifying numeral (or vice versa). For example, when counting in Maltese, 2 is expressed as *tnejn*, but in construction with a noun, the word for 2 is *żewg*. Maltese is the only European language I have found with such a suppletive alternative counting numeral. The best known example of this sort occurs outside Europe, in Chinese, where there are also two quite different words for 2, depending on whether one is reciting the counting sequence or expressing a proposition about some collection of two objects. The second word in the standard counting sequence is *erh*, whereas the word meaning 2 used with nouns (and their accompanying classifiers) is *liang*.

Less extremely idiosyncratic are cases where a counting numeral is **phonologically similar** to the quantifying form, but **not predictable** from it by a rule applying to other forms. For example, in German, the counting numeral *eins*, '1', is similar to, but not precisely predictable from any of the quantifying forms *ein, eine, einen, eines, einer, einem.*

Sometimes the counting numeral is the same as a pro-form numeral; German *eins* happens to be identical to one of the available pro-forms for an indefinite singular noun phrase. Similarly in Hungarian, 2 is either *két* or *kettö*. The longer form is used in counting and when no quantified noun is present (i.e. as a pro-form for an indefinite noun phrase); the shorter form is used to quantify a noun. *két/kettö* is the only simple lexical numeral in Hungarian which shows such a difference. For Basque, Saltarelli (1988:252) gives both *bi* and *biga* as forms for 2, and notes 'The forms of the numerals ... do not change when used as attributes. The only restriction seems to hold for the variant *biga* of *bi* "two", which cannot be used attributively modifying a noun.' Other grammars of Basque do not mention this detail. A specifically counting numeral need not be identical to the corresponding pro-form numeral. In some varieties of German which use *zwo* ('2') in counting, the pro-form for 2 is still *zwei*, the same as the attributive form.

The least idiosyncratic case of a difference between counting and quantifying numerals is where a **regular process** applies to one to yield the other. For example, in Gaelic, both the counting numerals and forms used when no head noun is present have a prefixed *a* (which may trigger an epenthetic *h* if the numeral begins with a vowel). Examples: (counting) *a h-aon, a dhà, a trì, a ceithir* versus (attributive) *aon, dà, trì, ceithir*, '1, 2, 3, 4'. Tatevosov gives a particularly clear example of distinct counting numerals, from Godoberi, in this case formed by adding a suffix to the ordinal forms.

> Suffix -la- brings the meaning of performativity. As ordinal numerals in -la- are used to count things, in our notation we use COUNT to symbolize their meaning.
>
> Numerals like ce:łila, k'e:łila, łabułila etc. would have been completely inappropriate if they had been placed into (27) - (30) [examples of attributive and predicative use of ordinals, JRH].
>
> ...

Let us now imagine a boy counting the blocks:

(36) ce:-łi-la, k'e-łi-la, łabu-łi-la,
one-ORD-COUNT two-ORD-COUNT three-ORD-COUNT

First, second, third, ...
(Tatevosov, 1994:80-82)

Interestingly, the 'performative' counting numerals in Godoberi can be embedded into sentences by the use of a further attributive suffix, added after the *la*, COUNT, suffix.

The table below shows, for each number from 1 to 10, how many languages have a counting form that is distinct from any quantifying form for that number.

Number	Suppletive	Idiosyncratic variant	Regular variant	No distinction
'1'		2	2	13
'2'	1	4	2	12
'3'		1	2	14
'4'			3	14
'5'		1	2	14
'6'		1	2	14
'7'			3	14
'8'		1	2	14
'9'			3	14
'10'		1	2	14

Table 1: Distinct counting and quantifying numerals (sum across 17 languages – Alb, Arc, Bsq, Eng, Fin, Fr, Gdb, Grm, Grk, Heb, Hng, Ice, Lzg, Rmny, Mlt, ScGl, Sin.)

It was expected that the occurrence of distinct counting forms would be greatest for the number 1, and decrease sharply and steadily thereafter. In this sample, the distinct forms peak at 2, and the decrease thereafter is uneven. Perhaps the original expectation was wrong, because (although the counting sequence of course begins at 1) it is only the utterance of a form for 2 after a form for 1 that confirms that the activity of counting is taking place. This sample is, however, very small; a larger sample would give a more reliable and perhaps clearer result.

This section has dealt with differences between **counting** numerals used in the recited counting sequence (i.e. not in construction with nouns and not embedded in a sentence) and **quantifying** numerals used attributively in noun phrases. A third kind of numeral sometimes distinct, a pronominally-used form, has also been mentioned in passing.

Other differences, sometimes loosely called differences in 'counting systems', but in fact dependent on the semantic or syntactic class of the noun modified, will be discussed in section 3.1.2. Before leaving this topic, I will mention Archi sheep-counting numerals, which not only have the 'performative' properties of the counting numerals but also are determined by the semantic class of the entities counted. In Archi, according to A.Kibrik (personal communication), the special sheep-counting numerals are formed by adding a suffix *-an* to bare numeral roots. These sheep numerals are used only in the actual act of counting sheep, and are not accompanied by any sister noun (such as the word for 'sheep'). For counting other entities, the speaker recites the sequence of bare numeral roots, with no suffix.

2.3 Ordinals

Ordinal numerals, when they occur, are usually at least as adjectival in their behaviour as cardinal numerals, and indeed are typically, across all languages, clear adjectives, displaying such adjectival features as gender and case agreement and often (though not always) having the same word order in relation to the noun as an attributive adjective. Cardinals, on the other hand, are much more often distinct in some way(s) from adjectives. In this paper, ordinals will be discussed mainly in terms of their paradigmatic (derivational) relationship with cardinals; the syntagmatic relationships of ordinals with modified nouns, being in most instances essentially identical to the behaviour of modifying adjectives, will barely be discussed.

Not all languages have a separate series of ordinal numerals. And some languages only have (or in practice their speakers only use) ordinals for a limited low-valued subset of numbers. Maltese only has distinct ordinals up to '4th'. In Spanish, although ordinals for values higher than 20 are available, e.g. *centésimo*, '100th', they are rarely used, and a construction with a cardinal is used instead.

Distinct ordinals are derived from cardinals. Cardinals are never derived from ordinals. It is useful to distinguish between **regular** and **irregular** derivations of the ordinal.

A **regular** morphological process requires no special statement for any particular lexical item, although predictable phonological modifications may be involved. Examples of regular ordinal formation are French suffixation by -*ième*, and English suffixation by *-th*.

Irregularity occurs in various degrees. Some idiosyncratic modification of the cardinal stem may accompany an otherwise regular morphological process, as in English *five/fifth*. Or the ordinal may bear some unpredictable resemblance to the cardinal, as with English *three/third*. The extreme case is **suppletion**, which typically involves the very low numbers, 1, 2, and perhaps 3. Examples with 1/1st are: Greek *ena/protos*, Welsh *un/cyntaf*, Italian *uno/primo*, Finnish *yksi/ensimmäinen*, and English *one/first*. Examples with 2/2nd are Greek *djo/devteros*, Welsh *dau/ail*, Italian *due/secondo* and Finnish *kaksi/toinen*.

A language can show variation between regular and irregular ordinals, which may be semantically or stylistically conditioned, as with French '2nd' *second* vs *deuxième*, or Finnish '1st' *ensimmäinen* (suppletive) versus *yhdes* (regular). A language can also show such variation between alternative irregular forms, as with Latin *alter* 'second of two' and *secundus* 'second of more than two'[2]. (For the regularization of irregular and suppletive ordinals in compound constructions, see a later section.)

Number	Suppletive	Irregular	Regular	No distinct ordinal form
'1'	14		3	
'2'	6	3	8	
'3'		6	11	
'4'		3	14	
'5'		2	14	1
'6'		1	15	1
'7'		1	15	1
'8'		1	15	1
'9'		1	15	1
'10'			16	1

Table 2: Suppletive and irregular ordinals. 'Regular' here includes cases of minor irregularity produced by some slight phonological shortening or truncation, such as English *fifth*. Sample: 17 languages – Alb, Arc, Bsq, Eng, Fin, Fr, Grm, Gdb, Grk, Heb, Hng, Ice, Lzg, Rmny, Mlt, ScGl, Sin.)

[2]It is somewhat surprising to find *alter* in the possible expressions for '22nd', *alter et vicesimus* and *vicesimus alter* given that *alter* is commonly said to be used for the second of just two items.

This table clearly shows a correlation between arithmetic value and morphological regularity of ordinal formation. Nevertheless it should be noted that, although suppletive forms for '1st' are overwhelmingly predominant, they are not universal. The three languages in this sample which form all their ordinals by a regular process from cardinals are Archi (A.Kibrik, personal communication), Godoberi (Tatevosov, 1994:78-80) and Lezgian (Haspelmath, 1993) all from the Caucasian family.

3 Simple lexical numerals modifying nouns

The focus of this paper is on the interaction of numerals with nouns when both numeral and noun belong to the same minimal noun phrase. Illustrated below are some of the most common ways in which cardinal numerals combine with nouns in noun phrases.

[$_{NUM}$ hat]	[$_{N}$ artista]	(Hungarian)
6	acrobat$_{(\text{SING})}$	

[$_{NUM}$ piat']	[$_{N}$ stolov]	(Russian)
5	table$_{+\text{PLU}+\text{GEN}}$	

[$_{NUM}$ dwy]	[$_{PREP}$ o]	[$_{N}$ wragedd]	(Welsh)
2$_{+\text{FEM}}$	of	wives$_{(\text{PLU,FEM})}$	

[$_{N}$ minissu]	[$_{NUM}$ tun]	[$_{CLS}$ den]	(Sinhala)
men	3	people	

The Welsh and Sinhala examples involve extra words, beside the numeral and the noun, a preposition in Welsh, and a numeral classifier in Sinhala. Such multi-word constructions are relatively uncommon in Europe, and discussion of them in this paper is, correspondingly, restricted to the next section (3.1).

3.1 Multi-word numeral-noun constructions

In such multi-word constructions, the extra word combines either with the noun, in which case it is a preposition, or with the numeral, in which case it is a numeral classifier. Both possibilities can occur together, as in English *five head of cattle*, where *head* might reasonably be classed as an archaic classifier. The more such extra words are present, the less direct, or 'tight' is the syntactic relationship between the numeral and the noun, there being often the possibility of an analysis involving the embedding of one NP inside another, with numeral and noun then belonging to different minimal NPs.

3.1.1 Prepositional constructions

The (partitive-like) use of a preposition to express indefinite cardinal modification of a noun is rare in European languages, and not a typical way of expressing cardinality in human languages generally. An example is Welsh, in which there are two ways of saying, for example, *nine men*.

naw o ddynion
9 of men(PLU)

naw dyn
9 man(SING)

According to Thorne (1993:149), there is no semantic difference between such expressions. The prepositional form is normally used after the more 'nouny' numerals *mil* 'thousand' and *miliwn* 'million', as in *mil o ddynion* '1000 men'. In Biblical Welsh, at least, there can be gender agreement between a numeral and a noun in such a construction, as in *ddwy o wragedd* '2(FEM) of wives'. This does not necessarily mean that the numeral and the noun are in close construction, as agreement can apply over quite long distances. In Welsh, as in many other languages, a preposition is also used in a partitive-like construction with a following definite NP, as in *naw o'r merched*, glossed by the entirely parallel English '9 of the girls'. Such constructions are not cases of a numeral immediately attributive to a noun, all within the same minimal NP. In such constructions numerals may still agree in gender with the noun, as in German *eine von den Frauen*.

The desirability of maintaining the generalization that prepositions govern whole NPs points in the direction of analyzing instances such as the Welsh *naw o ddynion* '9 of men' as having the structure shown below.

```
          NP
         /  \
QUANT     PP
  |      /  \
 NUM  PREP  NP
  |     |    |
  |     |    N
  |     |    |
 naw    o  ddynion
```

What makes Welsh unusual is the extension of this kind of structure to indefinites and to low-valued numerals. Such constructions with numerals, prepositions and indefinite nouns, semantically equivalent to ordinary attributive numeral-noun structures, are relatively rare, in Europe at least. Where such prepositional constructions do occur, it tends to be with the most high-valued, nounier, numerals. This was already seen in the obligatoriness of the prepositional construction with Welsh *mil* and *miliwn*, noted above. And French provides more examples, as in *quatre millions de personnes*. (But *une dizaine de voitures* is not a relevant example, as *dizaine* is not a true cardinal numeral.) This paper will have nothing further to say about these prepositional constructions, as there is much else to focus on in the commoner cases of tightly attributive numeral-noun constructions. Constructions where the noun sister of a numeral takes the genitive or partitive case, though no preposition intervenes, as in Russian or Finnish, are in some ways similar to the Welsh construction, and these will be discussed further (see especially subsection 3.3.2.2).

3.1.2 Numeral classifier constructions

The central case of a numeral classifier construction involves just three constituents, the numeral, the classifier, and the noun. These constituents may occur in any order, except that the numeral and the classifier are (almost?) always adjacent, forming a middle-level constituent of the whole construction. For example:

liǎng	ge	péngyou
2	Cls	friends

'Two friends' (Mandarin)

Sometimes expressions like *forty head of cattle* are cited as evidence of a classifier construction in English. But it should be noted that the preposition seen in such expressions is not present in the central case of a classifier construction. *Forty head of cattle* is thus a marginal example of a classifier construction (or perhaps not a classifier construction at all). German has some expressions which qualify better as numeral classifier constructions. (The following examples are due to Frans Plank.)

drei	Glas	Bier
3	glass	beer

Note the absence of a preposition, the singular number on *Glas* and the fact that the sequence *drei Glas* can be separated in a sentence from the noun *Bier*, as in:

Bier	hat	er	drei	Glas	getrunken
beer	has	he	3	glass	drunk

The German-type construction also occurs in Hungarian (E.Moravcsik, personal communication, citing Beckwith (1992)).

True numeral classifier constructions are very rare in Europe. As far as I can ascertain, no European language has such a construction as its general method of combining an attributive numeral with a noun. This includes Daghestanian languages, none of which make use of classifiers in combining numerals with nouns (A.Kibrik, personal communication). Nichols (1992:213) writes " ...numeral classifiers are found with fair frequency all around the Pacific rim but almost nowhere else ...".

Scottish Gaelic has a series of numeral words, valued from 1 to 10, variously labelled 'collective numerals' and 'numerical nouns', which could sensibly be analyzed as exemplifying a true classifier construction. These numerals are used only (and not obligatorily) with nouns denoting humans. The normal cardinal and the corresponding 'personal' forms are given below in parallel, for comparison:

	CARDINAL	'PERSONAL'
1	aon	aonar
2	dà	dithis
3	tri	triùir
4	ceithir	ceathrar
5	cóig	cóignear
6	sè, sia	sèanar
7	seachd	seachdnar
8	ochd	ochdnar
9	naoi	naoinear
10	deich	deichnear

Examples of the use of personal numerals are: *dithis bhan* 'two wives', *triùir chompanaich* 'three friends', *seachdnar dhaoine* 'seven men', *ochdnar chloinne* 'eight children' (Calder, 1923:128-129). It would seem possible to analyze the *n(e)ar* suffix as a numeral classifier bound to its numeral, though clearly there has to be allowance for morphological irregularity with the lower-valued numerals. Numeral classifiers can be fused to their numerals: Nichols (1992:132) mentions Gilyak and Nasioi as having such fused classifiers. The restriction to a semantic class (in the Scottish Gaelic case humans) is also typical of true classifiers.

Bulgarian has a similar series of numerals, applicable only to male humans. Scatton (1984:168-169) gives forms for 2 - 6 which are regularly formed from the core cardinals by suffixing *-ma*. An example is *dváma učeníci* 'two students'; **dvá učeníci* would be ill-formed. The suffix *-ma* functions exactly as a bound numeral classifier, restricted to a natural semantic class (male humans). (My Bulgarian informant, Ivan Derzhanski, also mentions similar, though marginal or archaic, forms for the numbers 7 - 10 and 100.)

Hungarian, too, has a series of numerals just for humans, with a suffix *-an* or *-en*, depending on vowel harmony. *négy* is '4', *négyen* is '4 people'; *tiz* is '10', *tizen* is '10 people'. While these Hungarian forms do resemble the Scottish Gaelic and Bulgarian cases just cited, there is an important difference which makes them less amenable to analysis as numerals with a classifier. The Hungarian 'human numerals' cannot occur with a sister noun. Thus **tizen ember* '4+HUMAN person' is ungrammatical (Anna Babarczy, informant). This distribution is not idiosyncratic, but results from the fact that these human numerals are clearly adverbial, as evidenced by a large number of other distributional facts (E.Moravcsik, personal communication).

Just outside Europe, numeral classifier constructions are common in Persian (many of the actual classifiers being borrowed forms of Arabic nouns, though

Arabic itself does not use numeral classifiers). The following examples are from Lazard (1957:91).

se nafar sarbâz
3 person soldier
'three soldiers'

panj tâ qali
5 piece carpet
'5 carpets'

3.2 Word order

The discussion in all (sub)sections from here on is largely restricted to tightly attributive numeral-noun constructions, where both numeral and noun are clearly within the same minimal NP. We continue now with instances where the numeral in such a construction is a single lexical item.

In discussing numeral-noun word order, we will concentrate almost entirely on unmarked constructions expressing the exact cardinality of some collection. Thus, reversals such as those which bring about a shift from cardinal meaning to ordinal meaning (as in English *one day* versus *day one*), or from exact meaning to approximate meaning (as in Russian) will not be considered.

A primary statistical generalization, holding across the world's languages, is that attributive numerals tend to precede nouns. But this generalization has an interesting wrinkle. Dryer, on the basis of a very large and representative sample, concludes

> ...the two orders of numeral and noun are equally common among OV languages. ...Outside of Africa, VO languages exhibit a strong tendency to be NumN: ...But in Africa there is a very strong tendency in the opposite direction, for the numeral to follow the noun in VO languages." (Dryer, 1992:118)

More usually, numeral-noun order is directly compared with adjective-noun order. Counts made by Greenberg, Hawkins, Rijkhoff and myself are summarized below[3].

Greenberg (1963)

	NA	AN
NumN	8	10
NNum	11	0

Hawkins (1983)

	NA	AN
NumN	32	47
NNum	39	0

[3]In arriving at the figures for these summary tables, I have eliminated certain cases of doubt, with as much consistency as possible. Thus, for example, from Rijkhoff's counts I eliminated cases with numeral classifiers, and cases where alternative equally marked orders seem to be possible. Greenberg's figures are from his (1963a:86). Hawkins's figures are summed over prepositional and postpositional languages (Hawkins, 1983:71-72,82). Rijkhoff's figures are gleaned from his (1993), tables 1 and 2. My own data are from Adg, Alb, Arc, Avr, Bsq, Blg, Eng, Fin, Fr, ScGl, Grm, Gdb, Grk, Heb, Hng, Ice, Kbr, Mlt, Rmny, Sho, and Sin.

Rijkhoff (1993) World

	NA	AN
NumN	3	11
NNum	5	2

Rijkhoff (1993) Europe

	NA	AN
NumN	13	68
NNum	2	0

Current sample

	NA	AN
NumN	3	11
NNum	5	2

Note first that no absolute universal emerges. Greenberg's and Hawkins' samples included no AN& NNum languages. Hawkins claims an exceptionless universal:

> We can therefore generalize Universal (VI) Prep $\supset$ (NNum $\supset$ NA) to cover both postpositional and prepositional languages, exceptionlessly:
>
> (VI') If a language has noun before numeral, then it has noun before adjective; i.e., NNum $\supset$ NA (equivalently AN $\supset$ NumN).
> (Hawkins, 1983:82)

In Rijkhoff's sample, two languages, Gude (Chadic) and Mangbetu (Central Sudanic) falsify this claim.

A search of Dryer's database (see Dryer, 1992) revealed 22 AN& NNum languages, from 9 language genera, concentrated in 3 parts of the world: (1) A large area of Africa, very roughly near Cameroon, (2) India and Nepal, and (3) Papua New Guinea. The relevant sample size was 492 languages[4].

From my own small sample, Sinhala falsifies Hawkins' claimed absolute universal, which nevertheless remains an impressive statistical universal. Relevant Sinhala examples are:

ge-val tun-ak
house-PLU 3-INDEF
'three houses' (NNum)

loku pot dek-ak
big book 2-INDEF
'two big books'

[4] My thanks to Matthew Dryer for this search and to him and Bill Croft for helpful comments on it.

kata: tune: story 3+GEN 'of three stories'	pot tunətə: book 3+DAT 'for three books'

This is the normal construction for inanimate nouns in Sinhala. It is widespread and productively applied to fresh loan words, as the following example shows:

telefo:nkɔ:l tuna-kə-tə satə tiha-i
telephone-call 3-INDEF-DAT cents 30-PRED
'three telephone calls cost thirty cents'

For animate nouns, Sinhala uses a classifier construction, as in:

minissu tun den-ek
men 3 Cls-INDEF
'three men' (N-NumCls)

These tables show a strong statistical preponderance of NumN languages across the world. According to Rijkhoff's second table, this preponderance is especially strong in European languages.

Greenberg's (1963a:86) figures for numeral-noun order count the order of numeral and classifier in numeral classifier languages, and it seems likely that Hawkins counted in similar fashion. If this way of counting is not correct, it wrongly inflates the number of NumN languages in Greenberg's sample by 2 or 3, and in Hawkins' sample by considerably more. That it may be incorrect to conflate numeral-classifier order with numeral-noun order may be seen from Sinhala, where both classifier and non-classifier constructions are available, one for animate nouns and the other for inanimates. For inanimates, the order is NNum; for animates the order is N-NumCls.

For the purpose of stating a universal relating numeral-noun order to adjective-noun order, it is particularly inappropriate to count the order of numeral and classifier, since one is then not comparing like with like. The normal understanding of such universal statements is that a term (such as 'Noun') refers to the same grammatical entity on both sides of the implication sign. With Greenberg's method of counting, the term 'Noun' can refer to a noun in the context of adjective-noun order, but to a classifier in the context of numeral-noun order. For example, in a Sinhala expression such as:

hundə lamai tun dena:tə
good children 3 Cls+DAT
'to the three good children'

lamai would be counted as the relevant noun, establishing AN order, but to establish NumN order, Greenberg would count the classifier *dena:* as the 'relevant noun'; this is clearly, then, not a comparison of word orders in relation to the same element. It might actually be better to separate numeral classifier languages from the rest in stating word order correlations, as the presence of a classifier indicates a clearly distinct principle for encoding cardinality in relation to counted objects. (See Dryer (1992:119-120) and Greenberg (1975) for some useful comments and data in relation to numeral classifiers and word order.)

With a small set of Sinhala nouns referring to frequently grouped objects (e.g. knees, parents), there is also optional, stylistically significant, NumN order for some numbers above 1[5].

The very lowest-valued numerals in a language sometimes have a different order from the rest. Two mixed 'num N num' languages from Europe are Basque and Maltese, in which the postnominal order only applies for the numeral for 1 (in Vizcayan Basque the numeral for 2 also, exceptionally, follows the noun). In my own sample, three of the NA languages (Basque, Hebrew and Maltese) place the numeral for 1, exceptionally, after the noun, like an adjective, and all other numerals precede the noun, and these languages were accordingly counted as NumN. Of the four NNum languages in my sample, the two European ones are Kabardian and Adyge, both Northwest Caucasian and NA, NNum. Here again the forms for '1' pattern idiosyncratically, preceding the noun while all the other numerals follow (Greenberg, 1989:107-108). The other two, non-European, NNum languages in the table are Shona (NA) and Sinhala (AN). In Shona, as in almost all Bantu languages, numerals behave exactly as adjectives. In Sinhala, clearly, most numerals behave neither like Sinhala adjectives nor like numerals in most other languages; but here again just the numeral for 1, exceptionally, may optionally occur before the noun, like a Sinhala adjective and like numerals in most other languages.

Together, the Basque, Hebrew, Maltese and Sinhala exceptions suggest the secondary generalization that where adjective-noun and numeral-noun orders differ, there is the possibility for the lowest-valued numerals, and especially that for 1, to pattern like adjectives. The Adyge and Kabardian cases are not counterexamples to this, because in Adyge and Kabardian, adjective-noun and numeral-noun orders do not differ. Nonetheless, these languages are somewhat surprising in this respect, as the form for '1' is idiosyncratically ordered in a way distinct from both adjectives and other numerals. These languages are thus counterexamples to the general claim that 'There is a widespread tendency for lower numerals to be more like adjectives' (Corbett, 1991:135) and '(1) Simple cardinal numerals fall between adjectives and nouns (2) if they vary in behaviour it is the higher which will be more noun-like.' (Corbett, 1978:368) Without doubt, Corbett's generalizations capture a very significant tendency in numeral behaviour, and the Adyge and Kabardian data hardly upset their general validity. Greenberg (1989:108,111) hypothesizes that the ordering of the form for '1' in these languages is the sole historical remnant of a previous AN, NumN order. The possible combinations for languages in which the form for '1' is idiosyncratically ordered are set out below, with some example languages mentioned.

NA, NuN, N1. (Basque, Hebrew, Maltese)

[5] My information for Sinhala comes from work with a native informant, Kadurugamuwe Nagita, and from grammars by Karunatillake (1992), Garusinghe (1962), Geiger (1900), and Gair (1970).

NA,	NNu,	1N.	(Adyge, Kabardian)
AN,	NNu,	1N.	(Sinhala – N1 also possible)
AN,	NuN,	N1.	(no examples found)

3.3 Morphological Interactions

Between the numeral and the noun in two-word numeral-noun constructions, there can be a variety of morphological interactions. Across languages, the arithmetic value of the numeral correlates significantly with the degree of morphological 'activity'; the lowest-valued numerals tend to be the most highly inflected. These morphological processes operate in both directions in the surface string, either with the noun as source and the numeral as target, or vice-versa; in other cases, the source, or determinant, of the morphological process is a containing phrasal category, such as NP. And in some cases it is not clear which direction a process operates in. Numerals may be the targets of inflectional processes, in the sense that morphological structure is added to them, either by agreement with properties of a head noun (e.g. gender, case, number), or by a semantically or pragmatically motivated process, such as marking for (in)definiteness. And numerals may be the source of morphological processes, in the sense that they dictate the morphological structure of sister constituents, typically nouns. An outline of the main processes found is given in the table below. Later subsections of this section will expand on the details of the 'Definiteness', 'Case', 'Number' and 'Gender' lines of this table, with examples from languages.

Grammatical features involved in Numeral-Noun constructions.

FEATURE	PRIMARY TARGET	ORIGIN
Definiteness	Whole NP	Pragmatics/semantics
Case	Whole NP	Semantics or clause structure
	Noun	Numeral-Noun structure
Number	Whole NP	Semantics
	Numeral	Lexicon entry for Noun
Gender	Numeral	Lexicon entry for Noun

(In this table, the distinction between 'clause structure' and 'semantics' reflects the difference between cases signalling grammatical relations, such as subject and direct object, and cases signalling such meanings as spatial relationships. Also, where a noun is said to be the primary target of case-assignment in a numeral-noun construction, the source is said to be the whole numeral-noun construction, rather than just the numeral, because this is a matter of government within a specific structural configuration, rather than a matter of agreement with inherent features of the numeral.)

3.3.1 Definiteness

The source of (in-)definiteness in NPs is pragmatic, typically involving considerations such as whether the speaker presupposes the hearer can identify the referent of the NP concerned. The exact pragmatic effects may vary somewhat from language to language.

3.3.1.1 Definiteness marked on (end of) whole NP Syntactically, (in-)definiteness is most typically marked on the whole NP at its extremity, with the marker sometimes being cliticized onto a single (leftmost or rightmost) constituent; this constituent may or may not be a numeral. Examples are:

Basque (Definiteness marked, indefiniteness unmarked):

gizon bat-a	gizon bat	
man 1-DEF	man 1	
'the one man'	'one man'	

bi	gizon-a-k	bi	gizon	bi-a-k
2	man-DEF-PLU	2	man	2-DEF-PLU
'the two men'		'two men'		'the two' (PRO-Form)

(For an interaction between definiteness and number in Basque, see subsection 3.3.3.1.)

Sinhala (Definiteness unmarked, indefiniteness marked):

ekə	pota-k
1	book-INDEF

pot deka-k	pot deka
book 2-INDEF	book 2
'two books'	'the two books'

deka-k	deka
2-INDEF	2
'two'	'the two' (PRO-Forms)

3.3.1.2 Definiteness marked on numeral The feature of (in-)definiteness rarely reaches down into the internal structure of NPs, and there is even less often any specific interaction of definiteness with numerals, but this does occur in a few Balkan languages.

In Bulgarian, there is typically a single marker of definiteness per NP, and this may fall on the noun, or an adjective or a numeral.

knigi-te
books+DEF

'the books'

červeni-te knigi
red-DEF books
'the red books'

dve-te knigi
2+DEF books
'the two books'

In Albanian, there are separate definite and indefinite numerals for the numbers 1-4. The paradigms for '1' are given below.

Albanian: Definite versus indefinite numerals, 1-4.

	INDEFINITE		DEFINITE	
	Masc.	Fem.	Masc.	Fem.
NOM	një	një	njëri	njëra
GEN	njëri	njëre	njërit	njërës
DAT	njëri	njëre	njërit	njërës
ACC	një	një	njërin	njërën
ABL	njëri	njëre	njërit	njërës

Lambertz' (1959) grammar, from which this information comes, also mentions distinct definite and indefinite forms for the numerals 2-4, but many, though not all, of his examples do not involve morphological variants of the numerals themselves, but rather a definite particle preposed to the numeral. My own informant, Eranda Kabashi, a speaker of the Geg dialect of Kosovo, recognizes fewer morphological distinctions among the numerals than Lambertz, but nevertheless has some distinct forms for definites and indefinites, for example:

dy vajza
2 girls
'two girls'

të dyja vajzat
the 2+DEF girls+DEF
'the two girls'

In German, there is an interaction between case and definiteness (or perhaps better 'determinedness'). Where a determiner (which happens almost always to be semantically definite) precedes, the determiner is case-marked, and the numeral is uninflected, as in:

dieser drei Männer
these+GEN 3 men

'of these three men'

But where no determiner is present, in which case the NP is interpreted as an indefinite, the numerals for 2 and 3 show a genitive case ending:

dreier	Männer
3+GEN	men
'of three men'	

3.3.2 Case

A broad and rough distinction can be made between cases assigned on the basis of clause-level structure or meaning, and cases assigned on the basis of NP-level structure. Typically, nominative and accusative cases are determined by the role of an NP within a clause, whereas genitive case is typically determined within an NP. Within the NP itself, case marking may or may not arise out of the relations between a noun and an attributive numeral. Thus the origins of case marking as it affects attributive numerals are two-fold: (1) flowing 'downward' from the structure of larger constituents, or (2) arising specifically from numeral-noun interaction. In some languages numerals exhibit just one of these case-marking origins, while in others they exhibit both, sometimes in complex complementary distribution.

3.3.2.1 Case assigned to numerals by clause-level structure

Case marked on whole NP. In Hungarian, all numerals 1-10 inflect for nearly all 20-odd cases, but case (assigned 'topdown', on the basis of clause structure) is only marked on numerals when they are free-standing, without a noun head, i.e. as PRO-forms for whole NPs. (The same is true of case marking on adjectives.)

Case marked on both Noun and Numeral. In Finnish, case in numeral-noun structures is sometimes wholly assigned 'top-down' on the basis of clause structure, and sometimes arises in part from the numeral-noun interaction. The distribution of the two types of case assignment is determined by (1) the value of the numeral concerned, with the numeral for 1 allowing top-down case assignment to the noun for all cases, and the numeral agreeing in case, and (2) the particular case assigned from clause structure, with numerals from 2 upwards allowing top-down case assignment to the noun, and agreeing with the noun in case, just for non-nominative and non-accusative cases. Here are some examples.

Finnish 1 in all cases, and 2+ in non-Nominative/Accusative cases.

yksi	kenkä	yhtä	kenkää	yhdellä	kengellä
1+NOM	shoe+NOM	1+PART	shoe+PART	1+ADESS	shoe+ADESS
'one shoe'		'of one shoe'		'on one shoe'	

kahden	miehen
2+GEN	man+GEN

'of two men'

In Russian, very similarly to Finnish, the numeral for 1 agrees with its sister noun in all cases, and numerals for numbers from 2 upwards agree in case with their sister noun just when the case assigned to the whole NP is neither nominative nor accusative.

In Romany, case on attributive numerals is assigned from clause structure, by agreement with the head noun, but the case-marking on the numeral is rather 'half-hearted', not showing the full range of case inflection exhibited by the noun. In Romany, nouns are case-marked by a two-tier system, with the possibility of a general non-nominative-marking suffix being followed by an outer suffix indicating more specific case information. But numerals, though inflecting for case, only take an inner tier of suffixes, distinguishing only between nominative and non-nominative. Thus note that in the examples below, there are four distinct case forms of the noun for 'children', but only two distinct case forms for the agreeing attributive numeral.

Romany (also illustrating '**semi-**' Case marking on Numerals).

trin	čavore	trin-e	čavor-en
3	children	3+OBL	children+OBL(ACC)
trin-e	čavor-en-ge	trin-e	čavor-en-go
3+OBL	children+OBL+DAT	3+OBL	children+OBL+GEN

In Greek, case-marking on attributive numerals arises from clause structure, by agreement with a head noun.

Greek

pendakosion	ginekon	tesaron	ginekon
500+GEN	women+GEN	4+GEN	women+GEN
'of five hundred women'		'of four women'	

Many Greek numeral words inflect for case, but the inflecting ones do not form a continuous sequence.

In Albanian, according to Lambertz (1959:89), the numerals from 1 to 4 inflect for case (see the paradigm given in the section on Definiteness above), but there is disagreement or dialectal variation here, as Newmark (1982:251) writes "When used as determiners before nouns, numbers are not marked for the grammatical category of case. In all cases they accompany the noun without undergoing any changes in form."

Case marked on Numeral only. Sometimes case is assigned to an NP from clause-level structure, but is marked overtly only on an attributive modifier, which can be a Numeral. An example is:

German

dreier Männer
3+GEN men
of three men

The table below shows the maximum number of cases distinguished idiosyncratically on numerals, that is, instances where case-marking on numerals is not by a regular productive process which also applies to other parts of speech, such as adjectives or nouns. The numbers show, for each language, how many distinct rows there are in the case paradigm for each numeral.

Numeral	Alb	Grm	Grk	Ice	Rmny	Rus	ZD
'1'	3	4	3	4	2	6	2
'2'	3	2	1	4	2	4	1
'3'	1	2	2	4	2	4	1
'4'	3	1	2	4	2	4	1
'5'	1	1	1	1	2	3	1
'6'	1	1	1	1	2	3	1
'7'	1	1	1	1	1	3	1
'8'	1	1	1	1	1	3	1
'9'	1	1	1	1	1	3	1
'10'	1	1	1	1	2	3	1

Table 3: Maximum number of grammatical cases distinguished idiosyncratically on numerals. Thus '1' in a cell indicates **no** distinctions of case, '2' indicates a single distinction, and so on. This table does not count languages in which there is fully regular productive affixing of case morphemes which can also apply to non-numerals, e.g. in Hungarian. The Albanian is that of Lambertz (1959).

This table shows that there is an overall tendency for the numeral for 1 to distinguish more cases than the numeral for 2, and for the latter to distinguish more cases than the numeral for 3, and so on. But there are clearly exceptions to such a smooth progression, as in Albanian (according to Lambertz) and Greek.

3.3.2.2 Case assigned to Noun by Numeral-Noun structure In Russian, when the case assigned by higher clause structure to an NP is either nominative or accusative, the numerals for 2, 3 and 4 assign genitive singular case to a governed noun, while higher numerals assign genitive plural.

Russian Nominative and Accusative NPs.

dva	stolá	piat'	stolóv
2	table+GEN	5	tables+GEN
'two tables'		'five tables'	

In this aspect of its grammar, Finnish is quite similar to Russian. In Finnish, with numerals above 1, when the case assigned to the whole NP by clause structure is either nominative or accusative, this case is marked on the numeral, but the noun receives partitive case, which can be analyzed as assigned by the governing numeral.

Finnish Nominative/Accusative NPs.

kaksi	miestä
2+NOM	man+PART
'two men'	

To mention a spurious case-assignment by a numeral, Scottish Gaelic, *dà*, '2', is often said to assign dative singular case to its noun, but it is preferable to analyze this as a dual, which happens to have the same form as a dative singular.

3.3.3 Number

This section illustrates the independence of grammatical number (e.g. singular/dual/plural) from arithmetic value. The usual correlations are: the integer 1 with singular number, and integers above 2 or 3 with plural number. But, as will be seen, 1 may have, and be associated with, a plural form and higher integers may have, and be associated with, singular forms. The semantic choice of a particular arithmetic value determines the choice of a particular lexical item, a simple numeral. Thus, in English, the meaning '3' requires the word *three* and in Hungarian the word *három*. The lexical numeral chosen has grammatical properties of affecting, or being affected by, the grammatical number of related constituents.

3.3.3.1 Number assigned to noun by sister numeral The typical pattern, of course, is for the word meaning '1' to assign singular number to its sister noun, and for words denoting higher values to assign (dual or) plural number. This needs no illustration; in the present sample, this pattern is observed in the main attributive numeral-noun structures of Alb, Bsq, Blg, Fr, Grm, Grk, Heb, Ice, Mlt, Rmny, and Sin. Note that this statement applies only to simple lexical numerals up to 10; there are sometimes complications with higher valued, grammatically complex numerals, which we will mention in a later section.

A pattern observed in a substantial minority of languages has all numerals, including the simple lexical numerals up to 10, assigning singular number to sister nouns. In the present sample, this is true for Arc, Gdb, Hng, and Wls, three of which are geographically peripheral to Europe. (But see section 5.4 for languages which switch from plural to singular nouns after 10.)

In Basque, there is an interaction between definiteness and number, such that numerals from 2 upwards take singular nouns when there is no definiteness marker, but take plural nouns when a marker of definiteness is present. The plural marker is *-k* and goes outside the definiteness marker. (Houghton, 1961:23, and Karmele Rotaetxe, informant.)

hiru gizon
3 man
'three men'

hiru gizon-a-k
3 man-DEF-PLU
'the three men'

(Edith Moravcsik points out to me that this connection betwen definiteness and number is generally true for the singular-plural distinction, not only relative to numerals: definites are more likely to show number distinctions. See Biermann, 1982: 232,235.)

In Scottish Gaelic, *dà* '2' is sometimes said (e.g. by Nicolson, 1936) to take a dative singular noun. It makes more sense to analyze the so-called 'dative singular' as a dual, as several grammars (e.g. Reid and MacLeod (1935), Calder (1923)) do. There are in fact still a handful of nouns, all feminine, which have a distinct dual form; these are, according to an informant (Rob Mulally), falling into disuse, perhaps as a symptom of general language death.

Genuine nominal duals are rare in Europe. Slovene, Sorbian and Old Church Slavic have dual verbal endings, which do not concern us here. In Old Church Slavic, according to Corbett (1983:235), " ...with 2 the noun stands in the dual, the numeral agrees with it ... ".

In Maltese, a remnant dual suffix *-ejn* is still seen on some nouns. Relevant examples are: *xahar* 'a month', *xahrejn* '2 months', *xuhur* '3 or more months'. But this grammatical duality is not assigned by a sister numeral, as may be seen from such examples as the following: *żewg xuhur*, '2 months' contrasting with the ungrammatical **żewg xahrejn*, and *erbgha riglejn* '4 legs' (e.g. of a horse). In Maltese, the form for '2' assigns plural number to its sister noun.

In Russian, the numerals for 2, 3, and 4 assign genitive singular to a sister noun, whereas the numerals for 5 - 10 assign genitive plural. Although synchronically it seems reasonable to analyze this as a genuine singular, this modern singular is the remnant of a historic paucal.

3.3.3.2 Number assigned to numeral by sister noun Reversing the normal pattern of a numeral assigning grammatical number to a sister noun is the unusual case of pluralia tantum in some Slavic languages, including Russian and Bulgarian. A Russian example is *odni sani*, '1+PLU sledge(PLU) '. Russian thus has both singular and plural forms for the number '1', selected by idiosyncratic features of the sister noun. Similarly, Bulgarian has *ední kléšti* 'one [pair of] pliers', with a plural form of the numeral for 1 agreeing in number with a pluralia tantum noun. This is clearly a somewhat marginal phenomenon, in the languages themselves, and is seldom, if ever, found in other languages.

3.3.3.3 Meaning of whole NP determines number of numeral and noun

Finnish singular and plural numerals for all numbers. The rest of this section on grammatical number discusses the data from Finnish[6] which illustrate an unusual interaction between numerals and nouns. Something like it also occurs in Icelandic, but only for the first four numerals, whereas in Finnish (and also to a large extent Estonian) it applies to all the numerals, both simple and compound. The Finnish/Estonian phenomenon of a widespread distinction between grammatically plural and grammatically singular numerals is relatively uncommon, and barely mentioned in the literature on numerals and grammatical number, and so it is dealt with at some length here.

Morphology and syntax Most Finnish simple and complex numerals have both singular and plural forms, as do many of the quantifiers for 'many', 'few', etc. A selection is given below (where the examples taken are in the Nominative/Accusative case).

VALUE	SINGULAR	PLURAL
1	yksi	yhdet
2	kaksi	kahdet
3	kolme	kolmet
4	neljä	neljät
5	viisi	viidet
6	kuusi	kuudet
7	seitsemän	seitsemät

[6] This account is based on information given by Henna Makkonen, a native speaker of Eastern Finnish, and Ulla Tuomarla, a native speaker of Standard Helsinki Finnish. The account is supplemented by information from the few grammars which mention the possibility of distinctive singular/plural marking on numerals. The examples have been kindly checked by Jouko Lindstedt.

8	kahdeksan	kahdeksat
9	yhdeksän	yhdeksät
10	kymmenen	kymmenet
50	viisikymmentä	viidetkymmenet
100	sata	sadat
1000	tuhat	tuhannet
100000	satatuhatta	sadattuhannet
1000000	miljoona	miljoonat
'pair, couple'	pari	parit
'a few'	muutama	muutamat
'many'	moni	monet
'several'	usea	useat
'few, not many'	harva	harvat

The plural forms of the last four quantifiers given above, *muutamat, monet, useat* and *harvat*, are not interpreted in the same way as the plural forms of the numerals from *yhdet* 'one+PLU' upwards. Many Finns have been taught that the difference between singular *moni* and plural *monet* corresponds to a difference between English *many a boy* and *many boys*, though many an English speaker would find it difficult to say exactly what the semantic difference between these English expressions actually is. These non-numeral quantifiers will not be considered further here.

The plural numerals take plural nouns, with the usual meaning of '*n* groups of'. Both singular and plural numeral-noun phrases occur in most cases, as illustrated below with the noun *kenkä* 'shoe'.

CASE	'one shoe'	'one group (typically a pair) of shoes'
Nominative	yksi kenkä	yhdet kengät
Accusative	yhden kengän	yhdet kengät
Partitive	yhtä kenkää	yksiä kenkiä
Genitive	yhden kengän	yksien kenkien
Inessive	yhdessä kengässä	yksissä kengissä
Elative	yhdestä kengästä	yksistä kengistä
Illative	yhteen kenkään	yksiin kenkiin
Adessive	yhdellä kengällä	yksillä kengillä
Ablative	yhdeltä kengältä	yksiltä kengiltä
Allative	yhdelle kengälle	yksille kengille
Instructive		yksin kengin
Comitative		yksine kenkine
Abessive	yhdettä kengättä	yksittä kengittä
Essive	yhtenä kenkänä	yksinä kenkinä
Translative	yhdeksi kengäksi	yksiksi kengiksi

Yksi 'one' is the only singular numeral which agrees with a head noun in all cases, including nominative and accusative. In these two cases, all other formally singular numerals govern a noun in the partitive singular. But in the other cases, there is agreement in case between singular numeral and singular head noun. Between a plural numeral and a plural head noun, there is always agreement in case.

Semantics of plural numerals. The most straightforward and uncontroversial examples[7] involve things which commonly come in pairs, such as:

kahdet	sukat	kolmet	sukat
2+PL	sock+PL	3+PL	sock+PL
'two pairs of socks'		'three pairs of socks'	
neljät	silmät	kahdet	kädet
4+PL	eye+PL	2+PL	hand+PL
'four pairs of eyes'		'two pairs of hands'	
kolmet	saappaat	kolme	saapasta
3+NOM+PL	boot+NOM+PL	3+NOM+SG	boot+PART+SG
'three pairs of boots'		'three individual boots'	
			(Jensen, 1934:47)

Whitney (1956:173) mentions that the plural of the numerals is used with the comitative and instructive cases, for which nouns have no singular form (see the gaps in the table listing cases above.) Thus:

huone	kolmine	ikkunoineen
room	3+COMIT+PL	window+COMIT+PL+POSS
'a room with its three windows'		

kaksin	käsin
2+INSTR+PL	hand+INSTR+PL
'with two hands'	

In addition, according to Whitney, "the plural of the numbers is used with those nouns which occur only in the plural".

kahdet	sakset	kahdet	kasvot
2+NOM+PL	scissor(s)+NOM+PL	2+NOM+PL	face+NOM+PL
'two pairs of scissors'		'two faces'	

(Some further comments on 'two faces' will be made below.) Jensen (1934:47) adds another example of pluralia tantum:

[7] All examples not attributed to grammars were elicited from informants.

kahdet häät
2+NOM+PL wedding+NOM+PL
'two weddings'

The examples mentioned so far are straightforward and uncontroversial, and might give the impression that the occurrence of plural numerals with plural numerals is restricted to a handful of idiosyncratically conventionalized instances. But the construction is more widespread and productive. Both informants were asked, separately, for their reactions to various plural numeral-noun sequences. Their reactions were consistent on three major points.

1. Such plural-numeral plus plural-noun sequences are clearly grammatical. The informants reported clear and immediate intuitions about the 'correctness' of such sequences.

2. Such sequences are often hard to contextualize. "It's correct, but I can't think when I would want to use it", was a common reaction. But discussion almost always revealed a context in which the informant agreed that the sequence in question could (or even would) be used appropriately.

3. There is some hesitation in specifying the exact meanings of such sequences, in particular between readings

   - where an exact number of groups of things is specified (e.g. '4 groups of apples'),
   - with an inexact number of groups, each of some exact cardinality (e.g. 'groups of 4 apples').

   Almost always, informants settled on the former type of reading, sometimes after a little thought.

Things which do not typically come in pairs can be involved:

neljät kupit
4+NOM+PL cup+NOM+PL
'four groups of cups' (Cf. *neljä kuppia* 'four cups'.)

viidet pyörät
5+PL wheel+PL
'five sets of wheels'

neljät oppilaat
4+PL pupil+PL
'four groups of pupils (e.g. classes) '

kahdet paperit
2+PL paper+PL
'three sets of papers'

viidet paperit
5+PL paper+PL
'five sets of papers'

These last examples could mean, according to one informant, papers in triplicate or quintuplicate for a meeting, or, according to another informant, also the set each participant gets for the meeting.

The next example drew somewhat different reactions from the informants.

ostin	kolme	autoa
bought+1st+SG	3+ACC+SG	car+PART+SG

'I bought three cars'

ostin	kolmet	autot
bought+1st+SG	3+ACC+PL	car+ACC+PL

'I bought three sets of cars'

One informant said this last sentence was "almost ungrammatical", while the other found she could contextualize it to a situation where a rich person or a buyer from a large company bought three lots of cars. This informant also said that *kolmet autot* could be used in a car-racing situation where the racing cars had bunched into three groups.

In another set of examples the informants reacted differently, but consistently, to plural numeral-noun phrases. Here one informant (spontaneously) drew diagrams indicating what she thought the crucial meaning differences were. The first sentence is:

oppilaat	saivat	kolme	kirjaa
pupils	got	3+ACC+SG	book+PART+SG

This sentence is ambiguous. On a collective reading, there is a set of just three books which the pupils, as a group, receive. On a distributive reading, each individual pupil receives a set of three books; in a special case of this latter interpretation, each pupil receives copies of the same three books as the other pupils. The distributive reading (not restricted to this special case) can be made explicit by using *jokainen oppilas* 'each pupil' as the subject of the sentence.

The second sentence is:

oppilaat	saivat	kolmet	kirjat
pupils	got	3+ACC+PL	book+ACC+PL

Here, for this informant, the special case of the distributive reading is forced; she drew a diagram in which four (presumably it could have been more or less than four) students each, identically, possess a set of books, labelled A, B, and C. It would seem to be a matter of three book **types** (e.g. *Animal Farm, Buddenbrooks* and *Clarissa*), with each pupil receiving one token of each type. The other informant's volunteered contextualization of this sentence was a situation in which a teacher has three variously sized groups of pupils and gives each group of pupils one pile of books; we don't know how many books are in each pile, but there are exactly three piles.

What is common to the interpretations suggested by both informants is the idea of three sets (alias types, piles) of books.

The pluralia tantum noun *kasvot* 'face(s)' was mentioned earlier as requiring a plural numeral. Interestingly, there is another word for 'face', which is not pluralia tantum, namely *naama*. With this word, it is possible to use the plural form to make the 'groups of' reading. Thus, of the examples below, the first two are paraphrases (at least as far as the number of faces is concerned), while the third example means something different.

kolmet	kasvot	kolme	naamaa
3+PL	face(PL)	3+SG	face+SG
'three faces'		'three faces'	

kolmet naamat
3+PL face+PL
'three groups of faces'

This last expression could be appropriate in talking about a cartoon film, where three groups of faces float onto the screen.

With mass nouns, there can be interpretations based on temporal groupings of events involving the stuff concerned (beer, coffee).

ostin kolme olutta
bought +1st+SG 3 +ACC+SG beer +PART+SG
'I bought three beers (glasses of beer)'

ostin kolmet oluet
bought +1st+SG 3 +ACC+PL beer +ACC+PL
'I bought beer in three sessions (perhaps more than one glass at each session)'

tilasimme kolmet oluet
ordered +1st+PL 3 +ACC+PL beer +ACC+PL
'We ordered beer three times, and each time for every one of us'

Kahvi/kahvit 'coffee(s)' behaves like *olut/oluet* 'beer(s)', as above. One informant notes these as "special cases", perhaps because these are normally mass nouns.

Other examples, though quickly judged grammatical, were not so easy to contextualize, and informants correspondingly showed a reluctance to assign meanings to them. Such examples included

kolmet	tähdet	seitsemät	koirat
3+PL	star+PL	7+PL	dog+PL

The difficulty here seems to be that dogs and stars are not naturally or not often experienced in neatly parcelled groups.

The higher valued numeral words *sata* '100', *tuhat* '1000', and *miljoona* '1000000', have plural forms, but the plurals (like their English counterparts) do not carry the precisely denumerated 'groups of' meanings.

sata	tähteä	sadat	tähdet
100+SG	star+PART+SG	100+PL	star+PL
'(exactly) one hundred stars'		'hundreds of stars'	

With thousands and millions there can be less practical sureness, and thus the following pairs could be more or less paraphrases of each other, with perhaps some greater emphasis on the numerosity of the stars on the plural examples.

tuhat	tähteä	tuhannet	tähdet
1000+SG	star+PART+SG	1000+PL	star+PL
	'a thousand (or thousands of) stars'		

miljoona	tähteä	miljoonat	tähdet
1000000+SG	star+PART+SG	1000000+PL	star+PL
	'a million (or millions of) stars'		

3.3.4 Gender (alias Noun Class)

It is not always easy to distinguish between gender systems and systems of noun classes which trigger numeral classifiers. The Bulgarian, Scottish Gaelic and Hungarian constructions mentioned earlier, in the section on classifier constructions (Section 3.1.2), each using a special affix on a numeral just with human nouns, could also have been analyzed as evidence of minor genders. In this section, I have taken as instances of gender the phenomena referred to as 'gender' by the largely traditional grammars of the languages concerned.

Gender marking on attributive numerals is always in agreement with the inherent gender of the sister noun. Examples from Zurich German (Weber, 1964:132,133) are:

zwee Mane 'two men'
2 men

zwoo Fraue
2 women

zwäi Chind
2 children

drei Mane
3 men

drüü Chind
3 children

Where gender is not regularly and productively marked on numerals by the affixation of gender morphemes which may also appear on other parts of speech, such as adjectives, idiosyncratic gender-marking is typically restricted to just the first few numerals, as the following table shows.

N	Alb_L	Alb_N	Blg	Fr	Grm	Grk	Ice	Mlt	Rus	ScGl	Wel	ZD
'1'	2	1	2	2	3	3	3	2	3	1	1	3
'2'	2	1	2	1	1	1	3	1	2	1	2	3
'3'	2	2	1	1	1	2	3	1	1	1	2	2
'4'	2	1	1	1	1	2	3	1	1	1	2	1
'5'	1	1	1	1	1	1	1	1	1	1	1	1
'6'	1	1	1	1	1	1	1	1	1	1	1	1
'7'	1	1	1	1	1	1	1	1	1	1	1	1
'8'	1	1	1	1	1	1	1	1	1	1	1	1
'9'	1	1	1	1	1	1	1	1	1	1	1	1
'10'	1	1	1	1	1	1	1	1	1	1	1	1

Table 4: Maximum number of grammatical genders distinguished idiosyncratically on numerals. This table does not count languages in which there is fully regular productive affixing of gender (or noun-class) morphemes which can also apply to non-numerals, e.g. Shona. Alb_L refers to the Albanian of Lambertz (1959); Alb_N refers to the Albanian of Newmark et al. (1982).

This table shows that there is an overall tendency for the numeral for 1 to distinguish more genders than the numeral for 2, and for the latter to distinguish more genders than the numeral for 3, and so on. But there are clearly exceptions to such a smooth progression, as in Albanian (according to Newmark et al.) and Icelandic.

Gender (Noun Class) determines Numeral-Noun construction In Sinhala, animate Nouns with attributive Numerals take a Classifier construction. Inanimate Nouns don't take a Classifier construction. See the Sinhala examples above, in section 3.2.

3.4 A note on 'uninflected'

Numerals happen to illustrate an analytic problem which is certainly found in many other subsystems of languages, though seldom remarked on. The problem is that there are two distinct senses in which a form may be said to be 'uninflected', with differing consequences for related syntactic and semantic rules.

In German, *ein*, '1', inflects for all four cases. The numerals for 2 and 3 *zwei* and *drei* have specifically genitive forms, *zweier* and *dreier* (occurring only in the absence of a preceding determiner). Of the numerals for 4-10, one would reasonably say that *vier, fünf, sechs, sieben* ... are 'uninflected'. But the lack of an inflection for (undetermined) genitive has different consequences from the lack of inflection for the other cases, as shown by the table below:

	1	2/3	4/5/6...
NOM	ein Mann	zwei/drei Männer	vier/fünf/sechs... Männer
ACC	einen Mann	zwei/drei Männer	vier/fünf/sechs... Männer
GEN	eines Mannes	zweier/dreier Männer	NO FORMS – CIRCUMLOCUTION NECESSARY
DAT	einem Mann	zwei/drei Männern	vier/fünf/sechs... Männern

The lack of a genitive for some numerals forces a German speaker to choose an alternative construction, in which a genitive is not required; so there are two ways of saying 'the story of three farmers', but only one way of saying 'the story of five farmers'.

die Geschichte	dreier	Bauer		die Geschichte	von	drei	Bauern
the story	3+GEN	farmers		the story	of	3	farmers+DAT
				die Geschichte	von	fünf	Bauern
				the story	of	5	farmers+DAT

The 'gap in the paradigm' or the need for circumlocution only arises for numerals not preceded by a determiner. With a determiner, neither *zwei* nor *drei* is inflected for genitive, or any other case, and no circumlocution is resorted to for the genitive (or any other) case.

die	Geschichte	der	drei	Bauer
the story		the+GEN	3 farmers	

die	Geschichte	der	fünf	Bauer
the story		the+GEN	5 farmers	

The case paradigms of numerals preceded by determiners are complete, with each cell for a given word (except for *ein*) filled by an identical uninflected form. The case paradigms for numerals without preceding determiners, on the other hand, are only complete for 1, 2 and 3, with gaps in the genitive row for the higher-valued words; apart from *ein*, the non-genitive entries for each such numeral are identical and uninflected.

4 Internal structure of complex numerals

4.1 Overall phrase structure and semantics

All languages with numeral systems use addition and multiplication as the principal semantic (arithmetic) operations for the construction of complex numerals. The detailed interaction of phrase structure and semantics in complex numerals across a wide range of languages is treated in Hurford (1975), and revisited in Hurford (1987). Here, only as much outline of such structure as is useful for the discussion of attributive numeral-noun structures will be given in this introductory subsection.

The number 230,567 is expressed in English as *two hundred and thirty thousand, five hundred and sixty seven.* The phrase structure of this expression (slightly simplified) is as in the diagram below.

```
                                NUM
                  ---------------|------------
                 PHR                        NUM
             -----|-------          ---------|---------
            NUM           M        PHR      CONJ     NUM
    ---------|-----       |      ---|---      |    ---|---
   PHR     CONJ  NUM      |     NUM     M     |   PHR   NUM
 ---|---     |    |       |      |      |     |    |     |
NUM    M     |   PHR      |      |      |     |    |     |
 |     |     |    |       |      |      |     |    |     |
two hundred and thirty thousand five hundred and sixty seven
 2    100          30     1000    5     100         60     7
```

The rules giving rise to such structures are argued for in Hurford (1975). Constituents of the category NUM(BER) are interpreted by addition; constituents of the category PHR(ASE) are interpreted by multiplication. 'M' (an abbreviation for 'multiplier') is a syntactic subcategory peculiar to numeral systems, and contains words such as *hundred, thousand* and *million* and their analogues in other languages; 'M's (or 'bases', as they may also be called) especially the higher-valued ones, behave in many respects like nouns.

Structures similar in their main respects to this are appropriate to the complex numerals of the great majority of languages. Complex numerals are indeed relatively similar in their internal structure across languages.

4.1.1 The Packing Strategy

A powerful generalization over structures such as (1), applying to the great majority of languages with very few exceptions, is the so-called 'Packing Strategy' (Hurford, 1975,1987). This is a universal constraint on numeral structures stipulating that the sister node of a NUM(BER) in such a structure must have the highest arithmetic value permitted by the numeral rules and lexicon of the language. It is convenient to illustrate the working of the Packing Strategy by showing an actual counterexample to it; isolated counterexamples do exist. The Packing Strategy in fact predicts that for any one number there will be only one well-formed numeral expression, and this is shown to be too strong by examples such as the alternative English expressions for 2100, namely *two thousand one hundred* and *twenty one hundred.* These have the structures shown below, and the Packing Strategy predicts that only that of *two thousand one hundred* is wellformed.

```
          NUM                          NUM
           |                    -------|------
          PHR                  PHR           NUM
       -----|----            ---|---          |
     NUM       M           NUM     M         PHR
   ---|---     |            |      |       ---|---
  PHR   NUM    |            |      |      NUM    M
   |     |     |            |      |       |     |
 twenty one hundred        two thousand one hundred
```

The arithmetic value of both expressions is 2100. The M in the left-hand structure has the value 100, and this is not the highest value possible (i.e. still below 2100) for an M permitted by the grammar, since the grammar also contains an M with value 1000, namely the word *thousand* which occurs in the rival structure on the right. The Packing Strategy insists that wellformed numeral structures 'pack' the highest-valued constituents in at the highest structural level. For detailed illustration, see Hurford (1975). The complex numerals of European languages conform, with a few exceptions, to the Packing Strategy.

Certainly, European languages do not behave in any particularly characteristic way in relation to this Strategy.

The Packing Strategy predicts dominance relations in complex numeral phrase structures. It says nothing about other properties in which the complex numeral expressions of languages vary. Such other properties will be surveyed across a range of mostly European languages below.

4.2 Multiplication

4.2.1 Decimal and vigesimal bases

Decimal (10-based) numeral systems are prevalent throughout Europe. Pure vigesimal (20-based) systems, which occur sporadically throughout the rest of the world (e.g. in Mixtec and Yoruba) are found in Europe only at its Western and Eastern extremities, in all four Celtic languages and in Basque, and in all the languages of the Northwest Caucasian group except Kabardian (Greenberg 1989:107). Lezgian numerals are vigesimal (Haspelmath, 1993). But Godoberi numerals are decimal (Tatevosov, 1994:62). Further inside Europe, and in different language families, there are modern elements of vigesimal systems or evidence of vigesimal usage in earlier times. Standard French has a mixed vigesimal/decimal system. Price (1992:463-469) discusses traces of vigesimal counting in Southern Italy and Sicily and in other spatially or temporally remote dialects of Romance. Danish 50, 60, 70, 80, and 90 show clear signs of a vigesimal rationale, although no overt morpheme denoting 20 is present. Hamp (1992:898) reconstructs Common Albanian as having vigesimal forms for 40, 60 and 80 and decimal forms for 30, 50, 70 and 90, although grammars of the modern language (Lambertz, 1959; Newmark et al., 1982) give only decimal forms for 60 and 80; the modern Albanian form for 40 transparently contains the form for 2, suggesting a vigesimal tendency. However, 'Over almost the whole of the Balto-Slavonic area, the numeral system is rigorously decimal' (Comrie 1992:721). Emmerick (1992:311) mentions a vigesimal system co-existing with a decimal system in Old Iranian (just outside Europe).

4.2.2 Word order in multiplicative structures

In the vast majority of languages, the 'M' or 'Base' word in a numeral PHRASE (a constituent interpreted by multiplication) is ordered like a noun modified by a numeral (Greenberg 1989:105). Thus, due to the predominant NumN order, the 'M' comes after its sister constituent, as (recursively) in English *three hundred thousand*. As far as I know, the only European languages which do not order the constituents of multiplicative structures in this way belong to the Northwest Caucasian group (Kabardian, Adyge – Greenberg, 1989:107). Even in Kabardian and Adyge, the MNum order does not apply to the whole numeral system, and Greenberg hypothesizes that these languages are in the middle of a diachronic switch from NumM to MNum, prompted by a prior switch from NumN to NNum. Outside Europe, an example of the multiplicative 'M' preceding its sister numeral is Shona, consistent with its NNum (and NA) order.

ma-kumi	ma-viri	ma-kumi	ma-tatu
VI-10	VI-2	VI-10	VI-3
'twenty'		'thirty'	
ma-zana	ma-viri	ma-zana	ma-tatu
VI-100	VI-2	VI-100	VI-3
'two hundred'		'three hundred'	

(VI is the noun-class marker.)

In Sinhala, although numerals follow nouns, multiplicative 'M's precede their sister constituent numerals. That is, Sinhala 'M's are not ordered like nouns with respect to numerals. (Sinhala is already unusual in being AN and NNum.) Relevant examples are:

ge-val tun-ak
house-PLU 3-INDEF
'three houses'

tun si:yə
3 100
'three hundred'

tun daha
3 1000
'three thousand'

In summary, the following combinations are possible:

NumN and NumM	A majority of languages and most European languages.
NumN and MNum	No examples known.
NNum and MNum	A minority of languages, e.g. Shona, Adyge (part), Kabardian (part), Yoruba
NNum and NumM	A few languages, e.g. Sinhala, Maori, Tahitian

4.2.3 1-Deletion

No language in the sample makes explicit the formula 1×10, but several allow, or even insist on, the presence of a form denoting '1' in a multiplicational relationship with the higher-valued Ms, for 100, 1000, and 100000. Generally speaking, the higher-valued Ms retain the explicit multiplier denoting '1', reflecting the more nouny properties of the higher-valued Ms. German exhibits the whole spectrum of possibilities, as shown below.

'10'	OBLIGATORY	(zehn	*ein zehn)
'100'	OPTIONAL	(hundert	ein hundert)
'1000'	OPTIONAL	(tausend	ein tausend)
'1000000'	IMPOSSIBLE	(*million	eine Million)

A sample of facts is summarized in the next table.

	OBLIGATORY	OPTIONAL	IMPOSSIBLE
with '10'	15		
with '100'	9	3	3
with '1000'	8	4	3
with '1000000'	4	2	9

Table 5: Languages' use of 1-Deletion. Data from 15 languages: Alb, Arc, Bsq, Blg, Eng, Fin, Fr, ScGl, Grm, Grk, Heb, Hng, Ice, Mlt, Rus.

4.2.4 Other noun-like behaviour of Ms

Certain higher-valued numerals, typically for 100, 1000, and 1,000,000, can themselves have noun-like characteristics, in particular carrying inherent gender, or taking genitive or partitive case assigned to them by lower-valued sister numerals.

Where gender is involved, these higher-valued numeral words do not agree in gender with a head noun, but rather may themselves impose gender agreement on lower-valued numeral words. A Welsh example is in the gender contrast between masculine *cant/gant/chant*, '100', and feminine *mil/fil* '1000', as shown below.

dau	gant	tri	chant	pedwar	cant
2+MASC	100	3+MASC	100	4+MASC	100
two hundred		three hundred		four hundred	
dwy	fil	tair	mil	pedair	mil
2+FEM	1000	3+FEM	1000	4 +FEM	1000
two thousand		three thousand		four thousand	

Similarly, Greek *ciljades* 'thousands' is feminine, while *ekatomirio* 'millions' is neuter, and these forms trigger gender agreement in lower-valued sister numerals, as in:

tris ciljades
3+FEM thousands(FEM)

tria ekatomirio
3+NEUT millions (FEM)

The noun-like behaviour of higher-valued numeral words is also seen in their taking a case (typically genitive or partitive) assigned by a lower-valued sister numeral. Thus in Finnish, '100' is *sata*, nominative, if that is what is required by clause-level structure; but '300' is *kolmesataa*, with the partitive form *sataa*, even though the whole expression may be assigned nominative case by clause-level structure. Similarly, '1000' is *tuhat* (in nominative case), while '3000' is *kolmetuhatta*, with the partitive form *tuhatta*, even though the whole expression may be assigned nominative case by clause-level structure. As with non-numeral nouns, the partitive is only assigned by a numeral when the whole expression receives nominative or accusative case from clause-level structure. In other cases, these higher-valued numerals agree in case with their lower-valued sister numerals.

4.3 Addition

Variation along three separate parameters is involved in the signalling of addition. Each parameter has one end which is typically used for rather low numbers, and an opposite end which is typically used for higher numbers. These parameters are presented and explained in separate subsections below. At the end of this section, a table is presented summarizing the combinations of these parameters in a sample of 18 languages.

4.3.1 Single word versus several words

Lower numbers (above 10) tend to be expressed by morphologically complex single words, interpreted by addition. Higher numbers tend to be expressed by sequences of words. For example, English *fourteen* is one word, while *forty six* is two words.

Relevant criteria for deciding on how many words an expression has are:

1. Stress: there is typically one stress per word. By this criterion, German *neunzehn*, 'nineteen', is one word, while *neunundzwanzig*, 'nine-and-twenty', is two words. In a phrase such as *neunzehn Minuten*, the numeral usually receives one stress, while in a phrase such as *neunundzwanzig Minuten*, the numeral usually receives two stresses.

2. Susceptibility of the parts to separate inflection: 'internal' inflection suggests a separate word. For example, the ordinal of Latin *undecim* '11' is *undecimus*, with a single suffix. But the ordinal for *tredecim* '13' is *tertius decimus*, with two markers of ordinality. This suggests that *undecim* is one word, while *tredecim* is two words. Likewise in Finnish, converting the nominative *kuusituhatta* '6000' into the inessive case gives *kuudessat-uhanessa*, with the suffix *-essa* occurring twice, suggesting that this is two words rather than one, despite the orthography.

3. Interruptibility: the possibility of a non-numeral element (typically the sister noun) coming between the parts of a numeral indicates a word boundary. In Scottish Gaelic, for instance, '13' is *trì deug*, 'three teen'. A modified noun comes between the constituents of such an expression, as in *trì fir dheug*, 'three men teen', 'thirteen men'.

4. Freedom of parts: the possibility of finding the constituent parts of a complex numeral in a wide range of other constructions, especially as unbound forms, suggests that the numeral consists of more than one word. For example, the fact that English *-teen* only occurs as a bound form in numerals (except for a few forms derived from such numerals, such as *teenager* and *teens*) suggests that such numerals are single words.

5. Orthography: conventional spelling may be adduced as supporting evidence for word boundaries, although it is not a very reliable criterion, as the German and Finnish examples above show.

6. Speaker intuition: in the absence of other evidence, I sometimes resorted to asking an informant whether a form was 'one word or two'. Sometimes a confident answer was given, though it is not clear whether the informant necessarily had an appropriate concept of 'word' in mind.

These criteria for wordhood are not specific to numerals but are the criteria generally used for establishing wordhood. Some of the criteria correspond to one-directional implications, while others are stronger, carrying bidirectional implications. As with any such set of criteria in linguistics, clear-cut conclusions cannot be reached in all cases, and a small residue of uncertain examples remains.

4.3.2 Absence or presence of an overt connective

Addition may be signalled by mere juxtaposition of a pair of numerals, or more explicitly by some overt connective, such as a conjunction or preposition. Mere juxtaposition tends to be used for lower numbers, and an explicit addition-denoting morpheme tends to be used for higher numbers. Examples are: German *neunzehn* '19' versus *neunundzwanzig* '29', literally 'nine and twenty'; also Welsh *deuddeg* '12' versus *tri ar ddeg* '13', literally 'three on ten'.

A language may use different addition-signalling morphemes in different parts of the number sequence. Welsh, for example, uses the preposition *ar* 'on' as far as 39, but *a(c)* 'and' above 40.

Sometimes, but not often, the presence or absence of a connective depends on the value of the lower conjunct. Compare, for example, English *three thousand, three hundred* with *three thousand and three*, and French *vingt et un*, 'twenty and one', with *vingt deux*, 'twenty two'. There is a slight tendency across languages for a word for '1', as opposed to other 'digits', to trigger the use of an overt connective, as in the French example just mentioned.

Not all languages follow the general tendency to use a connective for higher numbers and to omit it for lower numbers. Hungarian, unusually, has a connective in the numerals for 11 to 19 and 21 to 29, but builds higher-valued numerals by mere juxtaposition. In Hungarian, *tiz* is '10', *hàrom* is '3', *harminc* is '30', and *száz* is '100'; relevant additive forms are *tizenhárom*, '13', *harminchárom*, '33', and *százhárom*, '103'. *-en* is a particle denoting addition in the '-teen' forms.

The addition-denoting connective may be a bound form or a free form. English *and* may be judged to be a free form. In Archi, on the other hand, a suffix *-(t)or* converts a word to an 'additive form'. For example, '20' is *q'a*, and '100' is *baš*; *q'otor* is a form for '20' used just when some lower-valued numeral is added after it, and similarly *bošor* is a form for '100' used when followed by an additive lower numeral (A.Kibrik, personal communication).

4.3.3 Low-high versus high-low word order

There is a tendency for lower-valued additive numerals to have the lower-valued summand before the higher-valued one, and for the reverse order to apply to higher-valued additive numerals. The reversal is seen quite early in Italian, where '16' is *seidici* 'sixteen', but '17' is *diciasette* 'ten-to-seven'. I am not aware of any language in which the opposite switch, from high+low to low+high, occurs.

4.3.4 Combinations of addition-related features: summary

The three parameters described above, single/multiple-word, ± connective, and low-high/high-low, define a three-dimensional space, or cube, within which additive constructions may be located. All of the eight possible combinations occur in my data. The combinations, with the abbreviations to be used for them in the following table, are:

lh : Single word, no connective, low-high order: e.g. Latin *undecim* '11'.

l+h : Single word, connective present, low-high order: e.g. Bulgarian *dvanádeset* '12'.

hl : Single word, no connective, high-low order: e.g. Basque *hamabi* '12'.

l+h : Single word, connective present, high-low order: e.g. Italian *dicianove* '19'.

LH : Several words, no connective, low-high order: e.g. Scottish Gaelic *ceithir deug* '14'.

L+H : Several words, connective present, low-high order: e.g. German *drei und zwanzig* '23'.

HL : Several words, no connective, high-low order: e.g. English *twenty three.*

H+L : Several words, connective present, high-low order: e.g. Maltese *mija u tlieta*, '1003'.

In addition to the above combinations, it is necessary to mention the possibility of numbers above 10 being expressed monomorphemically. This is strictly the **lack** of a construction signalling addition. In the table below, forms such as English *eleven* and *twelve* are counted as monomorphemic. It is not always easy to decide whether a form is monomorphemic or a very irregular bi-morphemic form. In the present count, for instance, French *onze, douze, treize, quinze, seize* have been counted as monomorphemic; there is room for some alternative analysis here. The abbreviation for such monomorphemic forms in the table below is 'mon'.

The table below shows the distribution of these classes of constructions, in my sample, in relation to ranges of numbers.

	mon	lh	l+h	hl	h+l	LH	L+H	HL	H+L
'11'	4	5	2	1		1			4
'12'	4	5	2	1		1			4
'13'	1	6	2	3		1			4
'14'	1	6	2	3		1			4
'15'	1	7	2	2		1			4
'16'	1	6	2	3		1			4
'17'		6	2	4	1	1			4
'18'		6	2	5		1			4
'19'		6	2	4	1	1			4
'21'-'99'							2	7	9
>'100'								7	11

Table 6: Syntax of additive constructions, summed across 18 languages (Alb, Arc, Blg, Eng, Fin, Fr, ScGl, Grm, Gdb, Grk, Heb, Hng, Ice, It, Lzg, Mlt, Rmny, Sin). Where there was variation in the ranges 21-99 and >100, the most frequent pattern was counted.

4.4 Minor arithmetical operations

In various languages there are sporadic traces of arithmetic operations other than addition and multiplication, notably subtraction, overcounting, and division.

Subtraction is rare in numeral systems generally. Where it occurs in Europe, it only involves subtraction of 1 or 2. Latin *undeviginti* '19', literally 'one from twenty', and *duodeviginti*, '18' 'two from twenty', are familiar. In Finnish, the forms for 8 and 9 transparently contain the forms for 2 and 1, respectively, although the rest of the word is not recognizably cognate with the Finnish word for 10.

Overcounting is a nonstandard arithmetical operation found sporadically in numeral systems. It involves anticipating the next-higher round number, but counting towards it, rather than subtracting from it. In Europe, there is a trace of overcounting in Finnish, where the forms for 11 - 19 all contain the form *-toista*, the partitive of an ordinal numeral meaning '2nd'. This can be interpreted as 'of the second decade'. (In the table on methods of signalling addition, above, *toinen* was counted as a higher-valued summand, i.e. as a form for '10' homophonous with the form for '2nd'.) The Russian *devjanósto* '90', which contains *sto* '100', is a relic of overcounting.

Division occurs very rarely. In Europe, it occurs in Welsh *hanner cant* 'half hundred' '50'. It is not clear how well integrated this expression is into the grammar of attributive noun modifiers.

5 Complex numerals modifying nouns

5.1 Word order

Almost always, complex numerals appear on the same side of their sister noun as simple numerals. Thus in English, both *three* and *three thousand three hundred and thirty three* precede any noun to which they are in attribution. Rightward shifting, or extraposition, of heavy constituents seldom affects numerals.

In my sample, rightward shifting with complex numerals occurs only in Welsh and Scottish Gaelic. In these languages, the complex numeral is broken up, or interrupted, by the modified noun. Thus, one part of the numeral remains on the left of the noun, while the remainder appears on the right. Some Welsh examples are:

tri bachgen ar ddeg
3 boy on 10
'three boys'

un mlynedd ar ddeg ar hugain
1 year on 10 on 20
'thirty-one years'

There is a greater variety of forms mentioned in accounts of Scottish Gaelic. Not only does the order of summands internal to the complex numeral vary somewhat, with both high+low and low+high orders attested, but also the 'sister' noun can occur in a variety of positions inside a complex numeral, as well as to its right (as with a simple numeral). Here are some examples from Paterson (1968):

ceithir	fichead	fear	's	a naoi-deug	
4	20	man	plus	9-10	

'ninety-nine men'

naoi-deug	's	ceithir	fichead	fear	
9-10	plus	4	20	man	

'ninety-nine men'

tri	fir	dheug	's	da	fhichead
3	man	10	plus	2	20

'fifty-three men'

da	fhichead	fear	's	a tri-deug	
2	20	man	plus	3-10	

'fifty-three men'

The sister noun is never entirely to the left of the whole complex numeral.

5.2 Preservation versus regularization of irregular processes.

In complex numerals, the irregular processes affecting simple numerals are either preserved or regularized. English preserves the *one/first, two/second, three/third, five/fifth* irregularities in compound ordinals, as in *two hundred and first, fifty-second, ninety-third* etc. French regularizes the *un/premier* suppletive irregularity, in *vingt et unième*. **Vingt et premier* is not wellformed.

A language may be 'mixed' in the preserving/regularizing dimension. Such a language preserves some of its irregular ordinals in compound numerals, but allows irregularities in others to be overridden by regular processes. In Welsh, for example, we have (*bymtheg* = '15'):

NUMBER	CARDINAL	ORDINAL	
1	un	cyntaf	(Suppletive)
2	dau	ail	(Suppletive)
11	un ar ddeg	unfed ar ddeg	(Suppletion regularized)
16	un ar bymtheg	unfed ar bymtheg	(Suppletion regularized)
17	dau ar bymtheg	ail ar bymtheg	(Suppletion preserved)
21	un ar hugain	unfed ar hugain	(Suppletion regularized)
22	dau ar hugain	ail ar hugain	(Suppletion preserved)

This mixed practice in Welsh may be connected with the fact that *cyntaf* 'first', alone among ordinals, usually follows its head noun (like an adjective), whereas other simple ordinals precede the noun. With compound numerals, both cardinal and ordinal, the head noun is embedded in the center of the numeral expression. Thus:

y	peth	cyntaf		un	funud	ar	hugain	
the	thing	1st		1	minute	on	20	
'the first thing'				'twenty-one minutes'				
yr	ail	blentyn		yr	ail	waith	ar	bymtheg
the	2nd	child		the	2nd	time	on	15
'the second child'				'the seventeenth time'				

Presumably, **y cyntaf munud ar hugain* would violate the normal postnominal ordering of *cyntaf* and **y munud cyntaf ar hugain* would violate the normal placement of a head noun within the compound numeral. Hence the resort to *yr unfed munud ar hugain*, which has, possibly to its advantage in the eyes of native speakers, the regularization of an otherwise irregular form.

This discussion of the Welsh case shows that choices on the dimension of the preserving versus regularizing of lexical irregularities in compound ordinals are not necessarily choices on a basic typological parameter, but may be derived from other, possibly idiosyncratic, features of a language. On the other hand, it often seems that the choice of preserving or regularizing is a basic, underived fact about the language in question. It is as hard to explain English *twenty first* and *hundred and first* as falling out from other facts of English grammar as it is to explain the converse French constructions *vingt et unième* and *cent unième* as a consequence of other facts of French grammar.

It might be argued that contrasts such as *twenty-first* versus *vingt et unième* are to be accounted for by appropriate definitions of the **word** in each language, rather than by a language-specific choice of whether to regularize otherwise irregular forms. For instance it might be argued that French *vingt et un* is a single word, to which the ordinalization process applies; thus the question of using suppletive *premier* would not arise, because there is then no question of ordinalizing *un*. On the other hand, English *twenty-one* would have to be regarded as a sequence of two words, with the ordinalization process applying just to the second word.

Certainly, language-specific, often idiosyncratic, definitions of words do play a part in determining the units to which the ordinalization process applies, as shown by examples such as the following:

		CARDINAL	ORDINAL	
(Latin)	11	undecim	undecimus	ONE WORD
	12	duodecim	duodecimus	ONE WORD
	13	tredecim	tertius decimus	TWO WORDS
	14	quattuordecim	quartus decimus	TWO WORDS
(Welsh)	11	un ar ddeg	unfed ar ddeg	SEVERAL WORDS

12	deuddeg	deuddegfed	ONE WORD
13	tri ar ddeg	drydedd ar ddeg	SEVERAL WORDS
15	pymtheg	pymthegfed	ONE WORD

Despite such examples, it remains questionable whether it is correct to analyze French *vingt et un* as a single word, in order to account for the regularization in *vingt et unième*. Against such an analysis are: (a) the orthographic spacing, (b) the use of the conjunction *et*, normally used between whole words, phrases and clauses, and not elsewhere found as an intra-word morpheme, except in a very few idiomatic expressions, and (c) the possibility of a slight difference in stress pattern between *vingt et un* and comparable single words, such as *continent* or *mendiant*.

Accounting for regularization of ordinalization processes by idiosyncratic definitions of word boundaries is an option that is not always available, as in the case of Welsh *unfed* NOUN *ar hugain*, where it is clear, because of the intervening noun, that *un ar hugain* is not a single word; and in any case, the ordinal morpheme *-fed* is a suffix, and not an infix (which it would have to be regarded as here, if *un ar hugain* were taken to be a single word).

In summary, although idiosyncratically determined word boundaries do play a part in the ordinalization process, there does seem to be an independent, sporadic process of regularization of otherwise irregular forms in compounds. In fact, French does it elsewhere, too. The irregular 2nd person plural of *dire* is *dites*, but in compounds such as *contredire, dédire* and *interdire*, this is regularized to *contredisez, dédisez* and *interdisez*; likewise, the future tense of *voir* 'see' is irregular, but the future of the compound *prévoir* 'foresee' is regularized.

5.3 Complex numerals as local or global targets, or non-targets

Morphological marking for definiteness, case, number and gender can be imposed on a simple lexical numeral by a sister noun or by clause-level structure, as described in earlier sections. When the numeral sister of a noun is complex, languages adopt different strategies for extending, or not extending, such morphological marking to the complex numeral. If inflectional marking applies to complex numerals as targets, the processes may be either local or global. The local/global parameter relates to the internal structure of the complex numeral expression (rather than, for example, any structure in a nominal modified by the numeral).

5.3.1 Local targets

A process with a local target affects only one word, typically the lowest valued, in the compound numeral.

5.3.1.1 Ordinalization Examples with ordinalization are:

(English) two hundred and forty-<u>first</u>

(French) vingt et unième
20 and 1+ORD
'twenty-first'

(Welsh) unfed ar hugain
1+ORD on 20
'twenty-first'

Where ordinal formation in compound numerals is local, the constituent affected is typically:

- adjacent to the head noun, (in at least one possible word ordering)
- the lowest-valued additive constituent in the phrase structure of the compound numeral,
- the rightmost element of the numeral.

These distributional features often coincide, but not always. Compare the following (with orthographic spaces inserted in the German for clarity):

(German)	der ein	und	zwanzigste	Mann
	the 1	and	20+ORD	man
	'the twenty-first man'			
	der hundert	erste	Mann	
	the 100	1+ORD	man	
	'the hundred and first man'			
(Welsh)	yr unfed	waith	ar	ddeg
	the 1+ORD	time	on	10
	'the eleventh time'			
	y drydedd	salm	ar	hugain
	the 3+ORD	psalm	on	20
	'the twenty-third psalm'			

5.3.1.2 Definiteness Bulgarian provides an example involving definiteness. When numerals stand as definite PRO-forms, they take a cliticized definite article *-te*, as in *sedemté* 'the seven'. "In the case of compound cardinal numerals either the first or the last constituent may carry the article" (Scatton, 1984:171)

dvadesét i petté
20 + 5+DEF

dvadesetté i pét
20 + 5+DEF
'the twenty five'

5.3.1.3 Case Icelandic has local case assignment with complex numerals. When a case such as genitive or dative is assigned topdown from clause-level structure, only a final digit word in a complex numeral is inflected for this case, while any such words inside the numeral remain in the nominative, which must be regarded as the unmarked case. Examples (with orthographic spaces inserted for clarity) are:

tvö	hundruð	og	tveimur	mönnum
2+NOM	100	+	2+DAT	men+DAT

'to two hundred and two men'

þrjú	hundruð	og	þremur	mönnum
3+NOM	100	+	3+DAT	men+DAT

'to three hundred and three men'

5.3.1.4 Gender Gender agreement in Welsh numerals is also a process with a local target (but see below for a reservation on whether it is always possible to decide this); here just the lowest digit word (*dau/dwy, tri/tair, pedwar/pedair*) agrees in gender with the modified noun.

dwy	ferch	a	dau	gant
2+FEM	girl	+	2+MASC	100

'two hundred and two girls'

dau	fachgen	a	dau	gant
2+MASC	boy	+	2+MASC	100

'two hundred and two boys'

tair	ferch	a	thri	chant
3+FEM	girl	+	3+MASC	100

'three hundred and three girls'

tri	bachgen	a	thri	chant
3+MASC	boy	+	3+MASC	100

'three hundred and three boys'

5.3.2 Global targets

A process with a **global target** affects many or all of the words in the compound numeral. Global feature-spreading goes down both branches of a NUM structure, ending up on many words in a complex numeral. But this process can be intercepted by more local feature mechanisms within a complex numeral.

5.3.2.1 Ordinalization In Greek and Finnish ordinalization is global. In Latin ordinalization is largely global (the ordinals of '11' and '12' not being global). Examples are:

(Modern Greek)	ikostos	protos
	20+ORD	1+ORD
	'twenty-first'	

diakosiostos	triakostos	ektos
200+ORD	30+ORD	6+ORD

'two hundred and thirty-sixth'

(Latin)	tertius	decimus
	3+ORD	10+ORD
	'thirteenth'	
	septimus	decimus
	7+ORD	10+ORD
	'seventeenth'	

septimum	et	quinquagesimum	atque	centesimum
7+ORD	+	50+ORD	+	100+ORD
'one hundred and fifty seventh'				

(Finnish)	viides	sadas
	5+ORD	100+ORD
	'five hundredth'	

neljäs	kymmenes	yhdes
4+ORD	10+ORD	1+ORD
'forty-first'		

Global ordinalization is sporadic in Gaelic. MacFarlane (:178) gives these alternatives for 'the 167th man':

an	seachdamh	fear	ar	thri	fichead	is	ceud
the	7+ORD	man	+	3	20	+	100

an	ceud	's an	t-seachdamh	's an	tri	ficheadamh	fear
the	100	+	7+ORD	+	3	20+ORD	man

The second example here shows ordinalization on two words in the complex numeral, but not everywhere; *ceud* '100' is not ordinalized, as it might be, to *ceudamh*. One of two further examples given by MacFarlane (:179), also showing ordinalization on several, but not all, eligible words in a compound numeral is:

an	treas	mile	deug	's an	ceithir	cheud	's an	cóig	's an	tri	ficheadamh	fear
the	3+ORD	1000	10	+	4	100	+	5	+	3	20+ORD	man
'the 13,465th man'												

The label 'semi-global' is appropriate to ordinalization in such cases. The Gaelic examples seem very variable, even somewhat random, perhaps not surprisingly for such high and complex numerals.

Global ordinalization is less common than purely local ordinalization. It occurs in some members of a language family, but not in others, for example, in Finnish but not in Hungarian, sporadically in Gaelic but not in Welsh, in Icelandic but not in German, in Latin and Spanish but not in French or Italian.

5.3.2.2 Case Finnish, Russian and Greek case agreement in numerals is global. Here is an elaborate Finnish example in the inessive case, from Vainikka (1989:101):

miljoonassa	kahdessa	kymmenessä	viidessä	tuhanessa	kolmessa	sadassa	neljässa
1000000+INESS	2 +INESS	10+INESS	5+INESS	1000+INESS	3+INESS	100+INESS	4+INESS

'in one million, twenty-five thousand, three hundred and four'

After giving a similar example, Whitney writes:

> ...the larger numbers, too, take the case-endings in all their elements. But because the numbers are so cumbrous, in large numbers it is the common practice to inflect only the last member. (Whitney, 1956:173)

Likewise in Russian, where case inflection is global, speakers experience performance difficulties in keeping this up for very large numerals. Mayer (1978) describes an interesting experiment in which native Russian speakers were asked to read aloud sentences containing large numbers written in Arabic symbols. The sentences were constructed to elicit global case agreement in large numerals. Mayer's subjects deviated frequently, and rather unsystematically, from the prescriptive norm, especially with larger numerals. One of Mayer's examples was the number 50614 in a context requiring the dative case. The prescribed standard would be:

pjati-desjati	tysjačam	šesti-stam	četyrnadcati	rubljam
5+DAT,SG-10+DAT,SG	1000+DAT,PL	6+DAT,SG-100+DAT,PLU	14+DAT,SG	roubles+DAT,PL

For this number, only 7 responses out of 30 were correct according to the prescribed norm. Many responses contained most of the required dative numerals, but sporadically substituted the odd genitive or nominative.

5.3.2.3 Number Finnish number (singular/plural) agreement between numerals and head nouns, discussed in an earlier section, is also global, as shown by the following elicited example.

kahdet	kymmenet	kahdet	sukat
2+PLU	10+PLU	2+PLU	sock+PLU

"twenty-two pairs of socks"

The difficulty of finding semantic interpretations for higher-valued plural numerals makes it hard to find plausible examples of complex plural numerals; but this example seems to show that where such numerals can be found, the plurality feature is indeed spread globally throughout the complex numeral.

5.3.2.4 Gender Greek and Spanish gender agreement are also global-target processes. Greek examples are:

pendakosia	ikosi	tesara	pedia
500+NEUT	20	4+NEUT	children(NEUT)

'five hundred and twenty-four children'

pendakosies	ikosi	tesaris	ginekes
500+FEM	20	4+FEM	woman(FEM)

'five hundred and twenty-four women'

Spanish examples are:

tresciéntos	un	hombres
300+MASC	1+MASC	men(MASC)

tresciéntas	una	mujeres
300+FEM	1+FEM	women(FEM)

5.3.3 Natural constraints on globality

The intuitive idea of a global process is that it applies across the board, affecting every word within its scope (here assumed to be the whole of a compound numeral). The global processes just described, where a morphological feature permeates, or even 'saturates', a whole complex numeral, are typically limited by other characteristics of the language concerned. Such limits may be due to the limited morphological resources in the language, or to the supervention, within the complex numeral, of overriding processes triggered by nouny elements in the numeral.

5.3.3.1 Limited resources The fewer declinable forms there are in a compound numeral, the harder it is to tell whether agreement applies globally. Likewise, if morphological ordinalization only applies to a few numeral words, there are not many compound expressions which could testify to global ordinalization. In Maltese, for example, only 1, 2, 3 and 4 have distinct ordinals, and so, even if ordinalization applied to compound numerals (which it doesn't), an expression for, say, 5,555th or 6,666th could not possibly reveal global ordinalization. Ordinalization, too, may simply not apply to compound numerals. In some dialects and styles of Spanish, for example, "The ordinal numbers above ten are generally replaced by cardinals" (Duff, 1965:55).

Greek ordinalization seems to be only partially global, since in examples such as the following, ordinal marking does not penetrate as far as the word *ciljades* '1000'.

tris	ciljades	pendakosjosti	ikosti	tetarti	gineka
3	1000	500+ORD	20+ORD	4+ORD	woman

the three thousand five hundred and twenty-fourth woman

The failure of global ordinalization to apply to the whole complex numeral here can be attributed to the fact that Greek does not provide an ordinal for *tris ciljades*. In Modern Greek, a meaning such as 'the 3000th woman' has to be expressed periphrastically, as *gineka tris ciljades*, literally 'woman three hundred'. Ordinal forms such as *tris ciliostos*, given by some old grammars (e.g. Wied, 1910:90) are felt to be rare and very stilted by modern speakers.

Morphological constraints limit, in an obvious way, the words in a compound numeral which are affected by a process. If some of the words are indeclinable, or have no corresponding ordinal forms, any putatively global process affecting the whole numeral must 'skip' those words (or perhaps be said to 'apply vacuously' to them). This shows in the following Greek examples, where the indeclinable *ikosi* '20' is flanked by the inflecting *pendakosia* '500' and *tesara* '4':

pendakosia	ikosi	tesara	pedia
500+NEUT	20	4+NEUT	children(NEUT)

pendakosies	ikosi	tesaris	ginekes
500+FEM	20	4+FEM	children(FEM)

One would presumably wish to say that the Greek agreement here **is** global, and the fact that *ikosi* shows no agreement is not counterevidence to this claim. The simplest analysis attributes the lack of gender marking on *ikosi* to its basic morphological indeclinability, rather than to some syntactic constraint connected with its position in the compound numeral.

A limiting case, in which it is impossible to decide whether a process is local or global, is illustrated by French gender agreement. In French, just one numeral, *un/une* '1', inflects for gender. It inflects both as a simple lexical numeral and in compound numerals, as in *un homme, une femme, trente et un hommes, trente et une femmes*. But the syntax of French compound numerals ensures that, even in quite long numerals, there can only ever be one instance of *un/une* which could agree in gender with a non-numeral head noun. For instance, '1,101' is *mille cent un(e)*; '1,001,101' is *un million mille cent un(e)*, but here the masculine agreement on *un* is determined by the nominal character of *million*. One should beware of deciding arbitrarily, in the absence of crucial examples, that French gender agreement 'must be' local, or 'must be' global, because that (whichever one chooses) is the natural default assumption. As far as the French system is concerned, one simply cannot tell. Possibly, other languages might provide arguments for either local or global as the natural default.

The greater the number of numeral words to which morphological processes apply, the more words there are in a typical compound numeral for global processes to work on. Conversely, with relatively few words to work on, there will be relatively few crucial examples to distinguish between global and local processes. The globality of agreement and/or ordinalization, where they apply, varies along a scale of 'visibility' from clearly discernible to completely indiscernible. In general, globality, if present, is likely to be more visible in ordinalization than in agreement, because in languages generally there are more numerals with ordinal forms than with forms inflecting for gender and/or case.

5.3.3.2 Case assigned within a numeral Higher-valued numeral words, Ms such as *hundred, thousand* and *million* have many characteristics of nouns, as discussed above (sections 4.2.2, 4.2.3 and 4.2.4). In languages where a lower-valued sister numeral assigns case to an M, as to a noun, any case-assignment to the whole numeral filtering down from higher-level clause structure is blocked. One can compare Greek and Finnish in this respect.

In Finnish, where the noun phrase is in the nominative or accusative case, the numeral takes that case and the noun is in the partitive singular. In nominative or accusative instances of complex numerals, the more nominal numeral words *kymmenen* '10', *sata* '100', *tuhat* '1000', and *miljoona* '1000000' are, as nouns would be, in the partitive singular, while the other numeral words are in the nominative or accusative. This example (from Vainikka (1989:101)) shows this:

miljoona	kaksi	kymmentä	viisi	tuhatta	kolme	sataa	neljä
1000000+NOM	2 +NOM	10+PART	5+NOM	1000+PART	3+NOM	100+PART	4+NOM

'one million, twenty-five thousand, three hundred and four'

For other cases, both numeral and noun are in the (same) relevant case, and case agreement extends uniformly and globally throughout a whole complex numeral, as an earlier example, in the inessive case, showed.

In Greek, on the other hand, numerals do not assign case to a sister noun, and correspondingly top-down case assignment to a word such as *ciljades* '1000' is not blocked by processes internal to the numeral. For example:

tris	ciljades	pendakosia	ikosi	tesara	pedia
3+NOM	1000+NOM	500+NOM	20	4+NOM	children+NOM

'three thousand five hundred and twenty-four children'

trion	ciljadon	pendakosion	ikosi	tesaron	ginekon
3+GEN	1000+GEN	500+GEN	20	4+GEN	women+GEN

'of three thousand five hundred and twenty-four children'

(In these examples, only case has been glossed; gender is also relevant and will be discussed immediately.)

5.3.3.3 Gender assigned within a numeral Greek *ciljades* '1000' is a feminine noun, and blocks gender agreement in a complex numeral permeating topdown from a noun sister of the whole complex numeral.

tris	ciljades	pendakosia	ikosi	tesara	pedia

3+FEM	1000(FEM)	500+NEUT	20	4+NEUT	children(NEUT)

'three thousand five hundred and twenty-four children'

5.3.4 Coincidence of global processes in complex numerals

In a language where two morphological features, for example ordinality and case, are both assigned globally, there may be global double marking. Thus, in Finnish, more or less as in Greek, ordinality is expressed on all the component words of a compound numeral, and case agreement is also superimposed on all these ordinals. Whitney (1956:175) gives this example:

neljäs	tuhannes	viides	sadas	kuudes	kymmenes	yhdeksäs
4+ORD+NOM	1000+ORD+NOM	5+ORD+NOM	100+ORD+NOM	6+ORD+NOM	10+ORD+NOM	9+ORD+NOM

'four thousand five hundred and sixty ninth'

neljättä	tuhannetta	viidettä	sadatta	kuudetta	kymmenettä	yhdeksättä
4+ORD+PART	1000+ORD+PART	5+ORD+PART	100+ORD+PART	6+ORD+PART	10+ORD+PART	9+ORD+PART

'of the four thousand five hundred and sixty ninth'

Is there any correlation, within particular languages, between the various processes affecting complex numerals? For example, can one predict that if ordinalization is global in a language, then case-marking and/or gender-marking will also be global?

Prima facie, there appears to be a close correspondence between global ordinalization and global case and/or gender agreement patterns in compound numerals. A language which has one often has the other. The possible coincidence of global gender/case agreement with global ordinalization suggests that there may be a single structural property which numerals in a language may or may not have, determining their behaviour along the global/local parameter. Vainikka (1989) makes an interesting suggestion for a multidimensional representation of structure, designed to capture the coincidence of global behaviour in Finnish numeral agreement and ordinalization. It is attractive to speculate on the possibility of a systematic relationship. Unfortunately, however, any comparison of a language's ordinalization, case and gender agreement processes along the local/global dimension must be constrained by the natural limitations to global processes discussed above. For many processes, it will be difficult or even impossible, to tell whether they are, or are not, global. From the data in the present sample, no implicational hierarchies or predictions begin to emerge. One may say, at least tentatively, that all of the possible combinations seem likely to be found. Given below are tables of the logical possibilities, with languages classified (sometimes only tentatively) according to them.

ORDINALITY	GENDER	
global	global	Greek
global	local	Russian
local	global	?

local	local	French?

ORDINALITY	CASE	
global	global	Finnish
global	local	Icelandic?
local	global	
local	local	German?

Returning briefly to the preservation or regularization of suppletion in complex ordinals, there is clearly no correlation between this and the globality or locality of the ordinalization process. All four logically possible combinations occur.

- local, preserving irregularity (English, German)
- local, regularizing irregularity (French, Italian)
- global, preserving irregularity (Greek)
- global, regularizing irregularity (Finnish[8])

5.3.5 Complex numerals opting out of morphological processes

Sometimes a process affecting a simple numeral is not extended to complex numerals.

We have already mentioned the lack of ordinals for complex numerals in Maltese, and in some varieties of Spanish.

An example involving case is the nonexistence in German of genitive-marked complex numerals ending in a form of *zwei* or *drei*. While *dreier Männer* 'of three men' is grammatical, a corresponding complex genitive, such as the hypothetical **hundert dreier Männer* 'of a hundred and three men' is ungrammatical. To express this meaning, one has to resort to a circumlocution, such as *von hundert drei Männern*, which uses, not a genitive, but a dative case, and in which there is no case-marking on the numeral at all.

An example involving gender comes from Maltese, where, at least for one informant, gender marking is not carried through to the parts of a complex numeral. Relevant examples are:

ktieb	wiehed	karozza	wahda
book(MASC)	1+MASC	car(FEM)	1+FEM

wiehed	u	ghoxrin	karozzi	mija	u	wiehed	karozzi
1+MASC	+	20	cars(FEM)	100	+	1+MASC	cars(FEM)

[8] Not 100% : *kahdeskymmenes ensimmäinen* '21st' (preserving suppletion) is also used, though *kahdeskymmenes yhdes* (regularizing) is the standard.

5.4 Complex numerals influencing a sister noun

Simple lexical numerals may influence the case or the number of a sister noun (see sections 3.3.2.2 and 3.3.3.1 above, respectively). There are two main ways in which this influence can be taken up by a complex numeral. Sometimes the whole complex numeral, regardless of its constituents, assigns a property to a sister noun. And sometimes a single word in a complex numeral, typically the rightmost, determines a property of the sister noun.

5.4.1 Whole complex numeral influences sister noun

In French, for example, all complex numerals assign plural number to a sister noun, regardless of whether they end in *un(e)* 'one'. French speakers feel no unease at the juxtaposition of *un* with a plural noun in an expression such as *vingt et un chevaux* 'twenty one horses'. Nor, for that matter, does an English speaker feel any unease at the juxtaposition of *one* with plural *horses* in the corresponding English expression.

In both Maltese and Scottish Gaelic, but in no other European language, simple lexical numerals (above '1') assign plural number to their sister nouns, while complex numerals, i.e. those above '10', assign singular number to their sister nouns. In principle, the converse switch, from lexical numeral assigning singular and complex numeral assigning plural, might be found, but I know of no instance.

5.4.2 One word in complex numeral influences sister noun

By contrast, in Russian, a complex numeral ending in *odin/odna/odno* 'one', imposes singular number on its sister noun, as in *sto odin stol* literally 'hundred one table'.

Russian also provides an example where the rightmost word in a complex numeral determines the case of the sister noun. Where the complex numeral ends in *odin/odna/odno*, the sister noun takes whatever case is assigned to it by clause-level structure (and the numeral agrees). Where the numeral ends in *dva, tri* or *četyre* (2, 3, 4), the sister noun takes genitive (singular case). Otherwise the sister noun takes genitive plural case.

5.5 Residual conflicts and puzzles

We have described the interaction of case, gender and number in complex numerals. The assignments of these features do not necessarily mesh comfortably in a given language. There are instances, especially where case, gender and number are all involved, where a language simply does not resolve how to combine its various assignments of morphological features, and speakers are correspondingly often puzzled or embarrassed when asked how to express certain meanings. Typically the numbers involved are the sum of a round number and 1, such as 301, or 1001. I will mention two such cases, one from German and one from Greek.

In German, *ein*, '1', on its own is followed by a singular noun, and agrees in gender and case with its head noun (and is the only numeral word to agree in gender). But this agreement is not extended (at least for many speakers) to compound numerals containing *ein*. Thus we have *eine Frau* '1+FEM woman', but *hundert eins Frauen* '101 women', where the non-agreeing counting form *eins* is resorted to. Many German speakers feel somewhat uneasy with this latter expression, but almost all prefer it to any expression showing agreement, such as **hundert eine Frau(en)*. It turns out that there is no completely satisfactory way to say (for example) '101 (or 201, 301, ...) women' in German. In addition to *hundert eins Frauen*, Duden mentions, as an alternative, *hundert und eine Frau*, which inserts a conjunction *und* (not found in the corresponding counting expression), uses a singular noun, and has a gender-agreeing form of *ein*. But German speakers are often equally uneasy with this expression, sometimes expressing the judgement that it suggests a separation of the women into a group of 100, with one extra added. The problem seems to be an unease which is felt at the juxtaposition of a singular case-marked and gender-marked numeral word with a semantically plural noun determining (or attempting to determine) this case and gender marking. The strong binding between the two words indicated by the case and gender marking is violated by the conflict in grammatical number. This kind of conflict only seems to appear in languages with both gender and case agreement.

A similar problem exists in Greek. The case/gender paradigm for *enas* '1' is as follows:

	M	F	N
Nom	enas	mia	ena
Acc	ena	mia	ena
Gen	enos	mias	enos

The problem is most acute with feminine genitive forms, as *mias* is most distinctively different from the other forms. Asked to express 'of five hundred and one women', speakers vary in their judgements, but a common set of judgements finds the following text-book prescribed example quite ungrammatical:

pedakosion mias ginekon
500+GEN 1+FEM+GEN women(FEM)+GEN

Instead, several informants preferred a form in which the word for '1' agrees in either case or gender, but not in both, as in.

pedakosion mia ginekon
500+GEN 1+FEM+NOM/ACC women(FEM)+GEN

pedakosion enos ginekon
500+GEN 1+MASC/NEUT+GEN women(FEM)+GEN

And several speakers expressed an absolute preference, as sounding more natural, for a form in which the word for '1' agreed in neither case nor gender, as in:

pedakosion	ena	ginekon
500+GEN	1+NEUT+NOM/ACC	women(FEM)+GEN

Presumably *ena* is regarded as a neutral unmarked form, which is acceptable in instances of uncertainty. One informant's judgements were for complete gender/case agreement throughout, but she admitted that some of the resaulting forms sounded very awkward and she would normally try to find some circumlocution.

In Russian, which has both gender and case, a similar problem does not arise, because the sister noun is singular after a complex numeral ending in *odin/odna/odno*.

6 REFERENCES

Beckwith, Christopher (1992). "Classifiers in Hungarian", in Kenesei and Pleh, 197-206.

Biermann, Anna (1982). "Die grammatische Kategorie Numerus", in Seiler and Lehmann, 229-243.

Calder, George (1923). *A Gaelic grammar.* Glasgow: Alex MacLaren and Sons.

Comrie, Bernard (1992). "Balto-Slavonic", in Gvozdanović, 717-833.

Corbett, Greville G. (1978). 'Universals in the syntax of cardinal numerals', *Lingua*, 46:355-368.

Corbett, Greville G. (1983). *Hierarchies, targets and controllers: agreement patterns in slavic.* London: Croom Helm.

Corbett, Greville G. (1991). *Gender.* Cambridge: Cambridge University Press.

Corbett, Greville, Norman M. Fraser, and Scott McGlashan (eds) (1992). *Heads in grammatical theory.* Cambridge: Cambridge University Press.

Dryer, Matthew (1992). "The Greenbergian word order correlations", *Language* 68,1:81-138.

Duff, Charles (1965). *Spanish for adults* (2nd edition) London: English Universities Press.

Emmerick, Ronald (1992). "Iranian", in Gvozdanović, 289-345.

Gair, James R. (1970). *Colloquial Sinhalese clause structures.* The Hague: Mouton.

Garusinghe, Dayaratne (1962). *Sinhalese: The spoken idiom.* Munich: Max Hueber Verlag.

Geiger, Wilhelm (1900). *Litteratur und Sprache der Singhalesen.* Strassburg: Verlag von Karl Trübner.

Greenberg, Joseph H. (1963a). "Some universals of grammar with particular reference to the order of meaningful elements" In Greenberg 1963b:73-113.

Greenberg, Joseph H. (ed.) (1963b). *Universals of language*, Cambridge, Mass.: MIT Press.

Greenberg, Joseph H. (1975). "Dynamic aspects of word order in the numeral classifier", in Li, 27-45.

Greenberg, Joseph H. (1989). "The internal and external syntax of numerical expressions", in Kefer, Michel, and van der Auwera, Johan (eds.) *Universals of language*, (*Belgian Journal of Linguistics*, 4):105-118.

Gvozdanović, Jadranka (ed.) (1992). *Indo-European numerals*, (Trends in Linguistics, Studies and Monographs 57, ed. Werner Winter). Berlin: Mouton de Gruyter.

Hamp, Eric P. (1992). "Albanian", in Gvozdanović, 835-922.

Haspelmath, Martin (1993). *A grammar of Lezgian*, (Mouton Grammar Library 9). Berlin: Mouton de Gruyter.

Hawkins, John A. (1983). *Word order universals.* New York: Academic Press.

Houghton, Herbert Pierrepont (1961). *An Introduction to the Basque Language, Labourdin Dialect.* Leiden: E.J.Brill.

Hurford, James R. (1975). *The linguistic theory of numerals.* Cambridge: Cambridge University Press.

Hurford, James R. (1987). *Language and number.* Oxford: Basil Blackwell.

Jensen, Hans, (1934). *Finnische Grammatik: I Teil: Laut- und Formenlehre.* Glückstadt, Hamburg: Verlag J.J.Augustin.

Kenesei, Istvan and Csaba Pleh (eds) (1992). *Approaches to Hungarian, vol.iv, The structure of Hungarian.* Szeged, LATE.

Kibrik, Aleksandr E. (1994). (ed.) *Godoberi's noun phrase*, Eurotyp Working Papers, Theme 7: Noun Phrase Structure, Working Paper No.21, Konstanz.

Karunatillake, W. S. (1992). *An introduction to spoken Sinhala.* Colombo, Sri Lanka: M.D.Gunasena and Co.Ltd.

Lambertz, Max (1959). *Lehrgang des Albanischen: Teil III, Grammatik der Albanischen Sprache.* Halle (Saale): Max Niemeyer Verlag.

Lazard, Gilbert (1957). *Grammaire du persan contemporain.* Paris: Librairie C. Klincksieck.

Li, Charles (1975). (ed.) *Word order and word order change.* Austin, Texas: University of Texas Press.

Mayer, Gerald (1978). "The use of Russian numerals in oblique cases", *Revue Canadienne des Slavistes / Canadian Slavonic Papers* 20, 2:208-217.

Newmark, Leonard, Philip Hubbard, and Peter Prifti (1982). *Standard Albanian: A reference grammar for students.* Stanford, California: Stanford University Press.

Nichols, Johanna (1992). *Linguistic diversity in space and time.* Chicago: University of Chicago Press.

Nicolson, Alexander (1936). *Modern Gaelic: A basic grammar*, Archibald Sinclair. Glasgow: Celtic Press.

Paterson, John M. (1968). *Gaelic made easy: A gaelic guide for beginners*, (in 4 parts). Glasgow: Dionnasg gaidhlig na h-Alba (The Gaelic League of Scotland).

Price, Glanville (1992). "Romance" in Gvozdanović, 447-496.

Reid, Duncan (and Norman MacLeod) (1935). *Elementary course of Gaelic*. Glasgow: an Comuinn Gaidhealach.

Rijkhoff, Jan (1993). "Word order in the languages of Europe: The simple noun phrase (a preliminary report)", Eurotyp Working Papers ii/5.

Saltarelli, Mario (1988). *Basque* (Croom Helm Descriptive Grammars). London: Croom Helm.

Scatton Ernest A. (1984). *A reference grammar of modern Bulgarian*. Columbus, Ohio: Slavica Publishers, Inc.

Seiler, Hansjakob and Christian Lehmann (eds) (1982). *Apprehension: Teil I*. Tübingen: Günter Narr.

Tatevosov, Sergei G. (1994). "Numerical quantification in Godoberi", Chapter 6 of Kibrik (1994)., pp.59-84.

Thorne, David (1993). *A comprehensive Welsh grammar*. Oxford: Basil Blackwell.

Vainikka, Anne (1989). "Multidimensional case agreement" 11th Scandinavian Conference, Workshop on Finnish Syntax.

Weber, Albert (1964). *Zürichdeutsche Grammatik: Ein Wegweiser zur Guten Mundart*. Zürich: Schweizer Spiegel Verlag.

Whitney, Arthur H. (1956). *Teach yourself Finnish*. London: English Universities Press Limited.