Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Lingua 117 (2007) 751–757
Lingua
www.elsevier.com/locate/lingua

Linguistic perspectives on numerical expressions: Introduction

Norbert Corver[a,*], Jenny Doetjes[b], Joost Zwarts[a]

[a] *Utrecht University, Utrecht Institute of Linguistics OTS, Trans 10, Utrecht, The Netherlands*
[b] *Leiden University Centre for Linguistics, P.O. Box 9515, 2300 RA Leiden, The Netherlands*

Received 10 March 2005; received in revised form 10 March 2005; accepted 10 March 2006
Available online 30 June 2006

Abstract

How is our knowledge of the number system represented in numerical expressions in human language? After briefly discussing aspects of the development, morphosyntax, and use of number words, this introduction summarizes how the six contributions to this special issue approach this central question.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Numbers; Discrete infinity; Number words; Cardinals; Ordinals

1. Introduction

Knowledge of language and knowledge of the number system are two cognitive capacities that distinguish human beings from other animals. A core property which is shared by these two cognitive domains is that of discrete infinity: just like the series of numbers goes on indefinitely (you can always add one more), you can go on building linguistic structures by adding new linguistic material to the already built structure, as in *John and Peter and Sue and Betty and* . . . (Hauser et al., 2002). This property of discrete infinity accounts for the fact that there is no limit in principle to how many words a sentence may contain. In his *Language and Problems of Knowledge*, Chomsky (1998:169) speculates on the idea that the number faculty developed as a by-product of the language faculty. He states that "we might think of the human number faculty as essentially an "abstraction" from human language, preserving the mechanisms of discrete infinity and eliminating the other special features of language;" see also Devlin (2001). This thought-provoking idea is not incompatible with the observation that cultures exist that are not aware of the

* Corresponding author.
E-mail address: norbert.corver@let.uu.nl (N. Corver).

0024-3841/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.lingua.2006.03.006

possibility of counting and whose language does not contain a method for building indefinitely many number words (see for instance Everett, 2004). As Chomsky notes, the number capacity is present but latent; if placed in the appropriate environment (say, a technological society), a child from such a tribe could become an engineer or a physicist as readily as anyone else.

In those cultures where the ability of counting is manifest, language is used to refer to numerosities – where numerosity is a property of a stimulus that is defined by the number of discriminable elements it contains – and mathematical operations on numbers (as in *two and two makes four*). How language uses numerical expressions to refer to numbers has been studied both from both a cognitive point of view (Dehaene, 1997) and a linguistic point of view (Hurford, 1987). It is the latter perspective that concerns us here in this special issue.

For a linguist, the central question is how our knowledge of the number system is represented linguistically in human language. This general research question raises a number of sub-questions that we discuss in the next section, after which we summarize how the six contributions in this issue relate to these questions.

2. Research questions

2.1. The development of number words

One cluster of questions concerns the development of number words. A first question is how number words have come into existence. What are the technological, cognitive and linguistic resources that have gone into their development? Another question is how number words are acquired. Are they acquired in the same way as other lexical items or do children learn the number sequence as a list of meaningless speech forms? Could it be that children initially store fixed sequences of number words as complex units and later learn to form complex numbers from simple number words?

2.2. The structure of number words

A second cluster of question concerns the structure of number words. What are the structural rules for building complex number expressions? And what is the internal structure and grammatical behavior of complex numerical expressions (e.g. *twenty five*)? How is the variation that we see between languages constrained? Although in many languages the composite number forms a clearly syntactic unit, in certain languages the composite number is circumposed around the count noun (as in the Welsh noun phrase *tair ferch ar ddeg* three (fem.) girl and ten; ‘thirteen girls’) or mixed in other ways with syntactic patterns (like English *one out of seven*). The question obviously arises what syntactic mechanisms underlie such patterns.

2.3. Number words and the morphosyntactic system

A third class of questions is related to the distribution of number words in the bigger phrasal and morphosyntactic system, inside and outside the DP.

First, one may wonder what the lexical status of number words is. Are they lexical categories (e.g. N, A) or functional ones (Q(uantifier)) (Jackendoff, 1977; Giusti, 1991)? Should they rather be interpreted as hybrid (semi-lexical) categories in the sense that they display both functional and lexical characteristics (see Corver and van Riemsdijk, 2001)? Or should they really be treated as a separate word class, as in traditional grammars?

Second, there are languages that need insertion of a so-called numeral classifier in nominal projections modified by a numeral, as in the Chinese example *san-ben shu* 'three-classifier$_{\text{volume}}$ book'. This property typologically implies the lack of compulsory number marking on the noun (Greenberg, 1972). What is the syntactic and or semantic status of these numeral classifiers? The correlation between the absence of number and the presence of classifiers in numeral classifier languages becomes particularly interesting in view of languages such as Turkish and Hungarian, in which there is number marking on nouns, but crucially not in the context of numerals, as in: *három kutya/*kutyák* 'five dog/dogs'. Floating quantifiers are also interesting in this respect, as they quantify over entities that are introduced by an NP while they are located in the VP (*They have all three left* ≈ *All three of them have left*) (Sportiche, 1988). We would like to gain understanding in the variation that we see here. In English, for example, they are accompanied by *all*, in Spanish, they need a definite determiner (e.g. *los dos* 'the two'), while floating numerals in Japanese are followed by a classifier (e.g. *go-satsu* 'five-classifier$_{\text{book}}$') and do not need to be definite.

In relation to the morphosyntactic status of number words, it should be noted that the expression of number is not limited to definite cardinal numbers like *four* or *twenty four*. Number words are part of a bigger system of quantificational expressions that include indefinite number words like *many*, *few* and *several*. The question how such words relate to the number concept, and in what respects they are similar to or different from ordinary numerals sheds light on the status of the number words themselves.

2.4. The uses of number words

Finally, number words are not only used as cardinals (*four busses*), but also have other uses, as ordinals (*the fourth bus*) or nominals (*bus number four*) or in counting sequences (*one two three four*). These other uses have received far less attention in linguistic research than the cardinal use. As a result, there are still many open questions about the different uses, how they relate to one another and how they are formally expressed. What are, for instance, the linguistic means of deriving ordinals from number words?

3. Overview of the contributions to this volume

The articles in this special issue address the above questions on the basis of various linguistic expressions with the general aim to broaden our understanding of the linguistic expression of the number system.

In her contribution 'The Co-Evolution of Number Concepts and Counting Words' Heike Wiese sketches an evolutionary scenario for the development of numerals. Humans have a number concept that is not only used for cardinality (*five busses*), but also in ordinal number assignment (*the fifth bus*) and in uniquely labelling objects (*the #5 bus*). Numbers can fulfil this function because they provide an infinite progression of distinct entities, a number sequence, that can be used as a numerical tool. So, it is language itself that provides the numbers, by offering counting words as verbal numerical tools.

The human system of numbers is a linguistic system, that relates to another cognitive, but non-linguistic system that is also found in animals and infants and that is based on prenumerical concepts of quantity, rank and identity. The linguistic and the conceptual systems are related in humans, in a non-iconic way, by associating the relational structures of both systems, what Wiese calls dependent linking.

Wiese sketches an evolutionary scenario in which language opened the way to cognition because counting sequences based on verbal numerical tools and numbers evolved together. In the first stage cardinalities are represented through a kind of mental tallies that represent the size of the set iconically, but these tallies were also made visual through notches on sticks and body tallies (fingers). In the second stage a conventional order develops in the tallies, especially in the body parts used for counting, as seen in many cultures around the world. In this stage the *names* of these body parts can also end up in a conventional order. In the third stage an indexical link arises between cardinalities and the last item of a verbal counting list (something also noted in Hurford's paper). The last item has been shown to be cognitively more prominent. In this way, we get a link between elements in the counting sequence and cardinalities. In the fourth stage a counting sequence is born through the ritualized symbolic association between the relational structure of counting words and sets. This is also the stage where the recursiveness of the verbal tool provides constructive infinity.

By treating numbers as tools, Wiese draws away from a Platonic conception of numbers. Also important is the gradual evolutionary emergence of the number concept. Language plays a crucial role in this development, by providing the verbal material, the symbolic linking and the recursion. It is no accident that the species that has language also has a systematic number concept.

The contribution of James R. Hurford is called A performed practice explains a linguistic universal: counting gives the Packing Strategy.

Building on earlier work Hurford proposes a universal constraint on arithmetical combinations in compound numbers, the Packing Strategy (Hurford, 1975, 1987):

> The sister constituent of a number must have the highest possible value.

This Packing Strategy is meant to select one numeral expression from the set of expressions that simple phrase structure rules generate for a given number, constraining in this way massive overgeneration. It does this by comparing the semantic structure of alternative expressions. As a result, the number 2000 is expressed as *two thousand* (2×1000) instead of **twenty hundred* (20×100), because the first form does more 'packing', by taking 1000 instead of 100 as a base. Hurford also discusses counterexamples to the Packing Strategy, like the form *twenty one hundred* (21×100), which should not exist alongside *two thousand one hundred* ($2 \times 1000 + 100$), because *hundred* does not have the highest possible value and explains them from the salience of the counting sequence based on hundreds.

The paper proposes that two very general principles explain the Packing Strategy instead of the standardization hypothesis of earlier work. The Iterated Learning Model, a simulation of language diachrony, is used to test the hypothesis that the Packing Strategy arose as a gradual process of standardization. However, it turns out that this only works for numbers below 100, that do no require the more costly process of comparing alternative expressions. One practical constraint that is proposed as an explanation for the Packing Strategy is 'Go as far as you can with the resources you have.' If you have ten words in your counting sequence (*one . . . ten*), then this will make *ten and four* a natural expression, but not *seven and seven*. Notice that existing resources of pluralization of nouns (*tens*, *hundreds*) and conjunction (*ten and four*) are used to express complex numbers. The other principle 'Minimize the number of entities you are dealing with' explains why we say *three hundred thousand* ($(3 \times 100) \times 1000$) and not *three thousand hundred* ($(3 \times 1000) \times 100$). Using *thousand* minimizes the number of entities ('packages') to 300 instead of 3000. Hurford compares the strategy of packing to the way humans parse sentences. On the other hand, Hurford concludes that it is unlikely that the Packing Strategy would show up in many other domains, given that it derives from the very specific task of counting.

In their contribution 'On prepositional phrases inside numeral expressions in Polish' Pawel Rutkowski & Hanna Maliszewska discuss the syntax of a complex numeral construction in Polish, viz. NUMERAL + *na* 'out of' + NUMERAL (e.g. *trzy na sto Holenderek*, three.NOM on hundred.ACC Dutchwomen.GEN; 'three out of a hundred Dutch women'). A general background assumption of this analysis is that most Polish numerals belong to the category Q and head a functional projection QP, which is hierarchically located in between DP and NP. Besides these Q-numerals, Polish also has a small set of adjectival numerals (viz. 1–4) and nominal numerals (viz. equivalents of 'thousand', 'million', 'milliard', et cetera).

It is argued that the complex numeral construction NUM + na + NUM is made up of two separate DPs, one of which is embedded within a prepositional phrase that is headed by *na* 'out of' and adjoined to the 'main' DP. More in particular, the 'underlying' structure looks as follows:

(1) $[_{DP}$ $[_{QP}$ NUMERAL $[_{NP}$ Noun $[_{PP}$ na $[_{DP}$ $[_{QP}$ NUMERAL $[_{NP}$ Noun]]]]]]].

It is the first numeral of the sequence, i.e. the PP-external one, that heads the whole numerical expression and determines external syntactic relations of the entire noun phrase, such as agreement with the finite verb. It is further proposed that the surface manifestation of only one of the two nouns is due to PF ellipsis: one of the nouns is deleted under identity with the other. This may yield the following surface variants: (i) NUM + na + NUM + noun (involving deletion of the first, PP-external noun), and (ii) NUM + noun + na + NUM (involving deletion of the second, PP-internal noun).

It is, finally, argued that certain NUM + na + NUM-sequences (more in particular, structures of the type: A-NUM [na Q-NUM + noun]) have been – from a diachronic perspective – reanalyzed syntactically as a single complex 'mono-phrasal' numeral (i.e. [A-NUM na Q-NUM] + noun).

Whereas Rutkowski and Maliszewska investigate the structure of numerals inside the DP projection, Mana Kobuchi-Philip focuses on the structure of numerals in floated position in 'Floating Numerals and Floating Quantifiers'. The main question addressed in her paper is why numeral quantifiers can float in languages such as Japanese. In most other languages quantifier float is restricted to non-numeral quantifiers such as *all*, *each* and *both*. The contrast between the languages is illustrated in (2) and (3). *Both* and its Japanese counterpart *ryoohoo* can float in both languages, while floating of a numeral is only allowed in Japanese:

(2) a. The students [**both** walked].

b. gakusei-ga [**ryoohoo** aruita].
student-NOM both walked
'The students both walked.'

(3) a. *The students [**three** walked].

b. gakusei-ga [**san-nin** aruita].
student-NOM 3-CL walked
'Three students walked.'

Her explanation of this makes use of the fact that Japanese floating numerals contain a classifier which is part of the numeral. Floating quantifiers have two properties that determine their distribution. In the first place, the FQ has to contain a restriction. This restriction is part of the FQ, either as an inaudible nominal element or as a classifier. This part of the analysis goes back to earlier accounts of floating quantifiers and binominal *each* that did not take Japanese into account. In addition to this first requirement on quantifier float, Kobuchi argues that the floating quantifier has to be a QP containing nothing but a Q head. This means that the nominal restriction has to be part of this Q head. Kobuchi motivates this requirement as follows. If the nominal element defining the restriction of the FQ were an NP, it would need Case, and there is no Case available in the A-bar position occupied by the FQ.

The two requirements form the key to her answer to the question why Japanese has floating numeral quantifiers. The sentence in (3a) is out, because *two* does not contain a restriction. English numerals can never float, even in contexts where a classifier is present as illustrated in (4). This is explained by the second requirement on quantifier float. As *flocks* is a noun, it requires Case. Japanese classifiers, as in (3b) above, are part of the Q head and as such they do not need Case.

(4) *The birds have **three flocks** landed in the pumpkin field.

Further evidence for this approach comes from a comparison between Japanese and Chinese, another numeral classifier language. Even though both languages have numeral classifiers, only Japanese allows for floating numeral quantifiers. Furthermore, Chinese numeral classifiers are word-like, while the Japanese ones are incorporated in the classifier. Kobuchi connects these two properties and argues that Chinese classifiers are similar to the English ones, so that floating numeral quantifiers are ruled out because the classifiers need case, as does *flocks* in the English example in (4).

Richard Kayne ('*Several*, *Few*, and *Many*') makes a proposal introducing elements similar to numeral classifiers in English, a language that is usually not considered to be a numeral classifier language. He argues for the presence of an abstract noun NUMBER in the context of *many*, *few* and *several*. The paper starts out with the (old) observation that *few* and *many* are adjective-like in that they take comparative and superlative suffixes. However, in an example such as *few books*, what is modified by *few* is the number of books, not the books themselves. Kayne argues that this is so because the adjectives *few* and *many* modify a hidden noun NUMBER. Similarly, *little* and *much* modify the hidden noun AMOUNT. This claim can be reduced to an instance of a scattering principle: *few* and *many* cannot simultaneously express what is expressed by the adjectives *small* and *large* and what is expressed by the noun *number*, because UG imposes a maximum of one interpretable syntactic feature per lexical item. The presence of the hidden singular noun NUMBER is further motivated by the possibility of expressions such as *every few days*, where *every* seems to modify a plural (*days*). Adopting the silent NUMBER hypothesis put forward in the paper can explain this fact: *every* does not modify a plural here, but a singular (*few* $NUMBER_{sing}$ *days*).

Contrary to *few* and *many*, *several* and numerals cannot take comparative or superlative suffixes. But *several* and numerals are very different from each other as well. For instance, numerals can be used in ordinals, whereas *several* cannot (*the seventh/*severalth person*), and numerals can be used as indications of age, while this is excluded for *several* (*John is seven/*several*). Kayne argues that the analysis of *several* is, after all, quite close to that of *few*. While *a few books* corresponds roughly to *a small number of books*, containing the hidden noun NUMBER, *several books* corresponds to *more than a small number of books*. As such the hidden NUMBER analysis is extended to *several*, while the lack of degree modification of this expression is explained by the hypothesis that *several* is necessarily modified by a hidden degree modifier MORE THAN. This explains the differences between *several* and numerals: *the more than a small number-th person* and *John is more than a small number* being excluded as well.

Sjef Barbiers' article, called 'Indefinite numerals ONE and MANY and the cause of ordinal suppletion' starts from the observation that many languages have a suppletive instead of a regular derivational form for the ordinal FIRST (cf. Hurford, 1987; Veselinova, 1996). This also holds for Dutch, which has the word *eerst(e)* instead of *een-de* or *een-st(e)*. It is pointed out that the suffix *-ste* in *eer-ste* is a superlative suffix creating the superlative form of the old temporal adjective *eer* 'early'. The question is raised as to why the regular forms *een-de* and *een-ste* are excluded in Dutch (Similar effects are found in many other languages). Hurford (1987) and Veselinova (1996) have argued that this tendency to use a suppletive form for FIRST is a relic of earlier stages in the evolution of language in which languages did not have an ordinal system yet. For the expression of ordinal concepts, words would be used from a domain that is supposed to be cognitively more prior than ordinality, e.g. temporal adjectives or spatial prepositions. Barbiers rejects such an 'evolutionary' account and argues that the widespread absence of regular ordinal formation with FIRST is due to the fact that the cardinal ONE, as opposed to the other (i.e. plural) cardinals, is inherently indefinite and ordinal suffixes require inherently definite numerals. In other words, ONE does not have the right feature composition to combine with the (definite) ordinal suffix. For this reason, ordinal formation with ONE uses another formation 'strategy', viz. suppletion. As a first step in the explanation, Barbiers shows that ONE displays a morphosyntactic behavior that is quite similar to that of *veel* 'many', which also lacks the feature [+definite] and cannot have an ordinal suffix either. As regards the definiteness of the ordinal suffix, Barbiers argues that it refers to a (external) linear ordering. More in particular, he proposes that the ordinal suffix establishes a definite relation between a unique point on a linearly ordered sequence and the position of a person, which in the unmarked case is the deictic center of the speaker.

References

Chomsky, N., 1998. Language and Problems of Knowledge. MIT Press, Cambridge, Massachusetts.

Corver, N., van Riemsdijk, H., 2001. Semi-lexical Categories: The Function of Content Words and the Content of Function Words. Mouton de Gruyter, Berlin.

Dehaene, S., 1997. The Number Sense. Oxford University Press, Oxford.

Devlin, K., 2001. The Math Gene. How Mathematical Thinking Evolved and Why Numbers are Like Gossip. Basic books.

Everett, D.L., 2004. On the absence of numbers and numerals in Pirahã. Paper presented at the Workshop on Numerals in the World's Languages, Leipzig.

Giusti, G., 1991. The categorial status of quantified nominals. Linguistische Berichte 136, 438–454.

Greenberg, J., 1972. Numeral classifiers and substantival number: problems in the genesis of a linguistic type. Stanford papers on language universals 9, 1–39. Reprinted in: Makkai, A., et al., 1977. Linguistics at the crossroads. Liviana Editrice, Padova & Jupiter Press, Lake Bluff, Ilinois, pp. 276–300.

Hauser, M.D., Chomsky, N., Fitch, W.T., 2002. The faculty of language: what is it, who has it and how did it evolve? Science 298 (5598), 1569–1579.

Hurford, J.R., 1975. The Linguistic Theory of Numerals. Cambridge University Press, Cambridge.

Hurford, J.R., 1987. Language and Number: The Emergence of a Cognitive System. Basil Blackwell, Oxford.

Jackendoff, R., 1977. X' Syntax: A Study of Phrase Structure. MIT Press, Cambridge, MA.

Sportiche, D., 1988. A theory of floating quantifiers and its corollaries for constituent structure. Linguistic Inquiry 19, 425–449.

Veselinova, L., 1996. Suppletion in the derivation of ordinal numerals. A case study. Paper presented at the 8th Student Conference in Linguistics, New York, April 1996.