+ Models
ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT

ELSEVIER

Lingua xxx (2006) xxx–xxx

Lingua

www.elsevier.com/locate/lingua

A performed practice explains a linguistic universal: Counting gives the Packing Strategy

James R. Hurford

Linguistics and English Language, Language Evolution and Computation Research Unit, University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh, Scotland EH89LL, United Kingdom

Received 7 January 2005; received in revised form 4 August 2005; accepted 10 March 2006

Abstract

A strong constraint on the arithmetical combinations allowed in compound numerals, called the Packing Strategy, applies very widely to numeral systems across the world. A previous attempt to explain the existence of the strong universal constraint, in terms of a gradual socio-historical process of standardization, will not scale up to higher-valued numerals. It is proposed that the real explanation for the Packing Strategy is that it reflects two natural principles applied in the practical task of counting objects. These two principles, ''Go as far as you can with the resources you have'', and ''Minimize the number of entities you are dealing with'', are not specific to the counting task, but are of more general application to practical tasks.

1. (Re-)introducing the Packing Strategy

The ''Packing Strategy[1]'' is a universal constraint on numeral systems. It applies very widely to developed numeral systems, that is to any system which uses syntactic constructions to signal addition and multiplication. It is not completely without exceptions, so it is not a truism, but exceptions to it are rare. The Packing Strategy was developed within the conceptual framework of early generative grammar. This approach attempts to provide rules which generate, economically and with semantic correctness, all and only the well-formed expressions of a given language. Applied to numeral systems, this amounts to providing rules which generate all and

E-mail address: jim@ling.ed.ac.uk.

[1] The Packing Strategy was introduced and extensively illustrated in Hurford (1975), and further discussed in Hurford (1987). Despite being a strong universal constraint on numeral systems, it has not, to my knowledge, been discussed by other scholars interested in numerals.

0024-3841/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.lingua.2006.03.002

only the well-formed numeral expressions, accounting naturally for the fact that they denote exact large numbers. The goal is also, of course, to discover sets of rules which apply to all languages.

The Packing Strategy operates in conjunction with a small set of phrase structure rules, which are shared by all developed numeral systems. These rules are given below.

Universal Phrase Structure Rules:

$$\textbf{NUMBER} \rightarrow \left\{ \begin{array}{c} \textbf{DIGIT} \\ \\ \textbf{PHRASE (NUMBER)} \end{array} \right\} \quad \text{(Interpreted by addition)}$$

$$\textbf{PHRASE} \rightarrow \quad \textbf{(NUMBER)} \quad \textbf{M} \quad \text{(Interpreted by multiplication)}$$

(These rules follow the normal conventions for Phrase Structure rules: curly brackets indicate ‘either/or’ options – choose either the upper line or the lower; parentheses indicate ‘take-it-or-leave-it’ optional choices – thus a PHRASE consists either of a bare M or the sequence NUMBER M. The semantic interpretations, addition and multiplication, are only applicable where two constituents are chosen.)

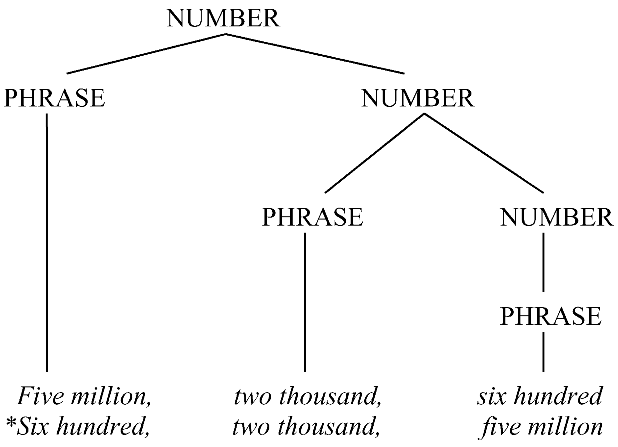
In these rules, DIGIT is the category of basic lexical numerals, such as English *one*, *two*, *three*, . . ., *nine*. M is the category of multiplicative base morphemes, such as English *–ty*, *hundred*, *thousand* and *million*. In fact a further rule can be postulated assigning internal structure to higher-valued Ms, but this will not be discussed here. It will be evident that these rules, being doubly recursive on the category NUMBER, massively overgenerate. The Packing Strategy is a global constraint that selects a single well-formed numeral for each numerical value, from among the many generated by the above phrase structure rules.

Here is an informal statement of the Packing Strategy, with some examples illustrating its working, in conjunction with these phrase structure rules.

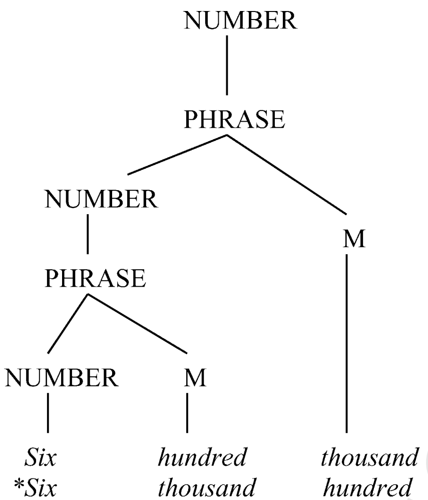
1.1. Packing Strategy: The sister constituent of a NUMBER must have the highest possible value

Here ‘highest’ means ‘highest equal or less than the total value of the whole expression’. And ‘possible’ means ‘within the set of structures also generated by the relevant phrase structure rules’. Thus, the Packing Strategy is a paradigmatic (or ‘global’) constraint, calling on comparisons between different potential forms from the lexicon and forms generated by the same set of rules. Some examples should make this clear.

In additive constructions, higher-valued constituents are packed into the structure nearer the top of the phrase structure tree. In most languages, as in English, this results in higher-valued constituents being ordered to the left of the structure. The Packing Strategy says nothing about linear order, but only about the hierarchical dominance relationships between constituents of numeral expressions; linear ordering of constituents is handled by the concomitant phrase structure rules above.



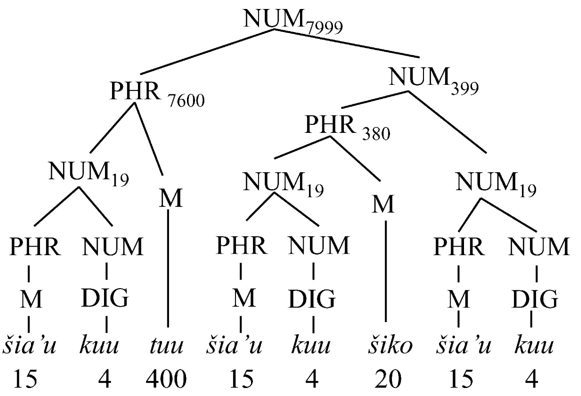
In multiplicative constructions, higher-valued multiplicative bases are packed nearer the top of the phrase structure tree. In English, as in many other languages, this results in the higher bases being ordered to the right of lower ones.



It will be evident from these examples that in complex cases the Packing Strategy can be envisaged as working cyclically from the top of a tree downwards, to get the highest numerals all the way to the top. In fact, a top-down algorithm is not absolutely necessary to implement the strategy, but the top-down idea can help understanding of the content of the constraint.

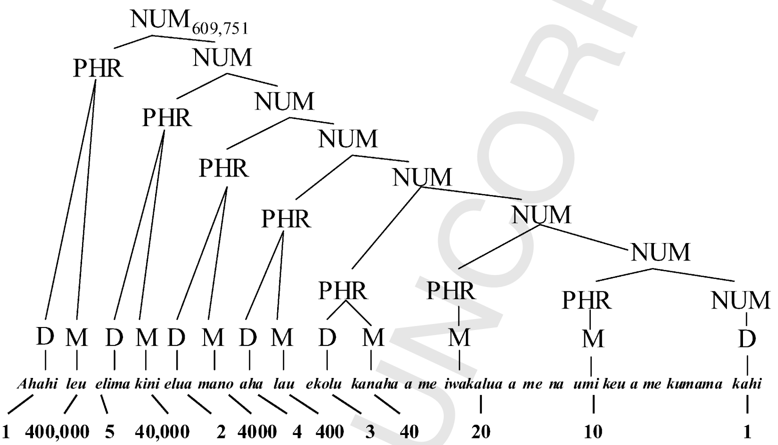
The irregularity of standard French numerals allows another insight into the working of the Packing Strategy. Standard French, in an idiosyncrasy which overrides the regular processes, does not express 70 as a multiple of 10. One would expect *septante* (which occurs in some non-standard dialects), but one finds *soixante dix*, 60 + 10. Why is 70 not expressed as some of the other apparent possibilities, such as **cinquante vingt*, 50 + 20, or **quarante trente*, 40 + 30? The Packing Strategy picks the right structure (given the overriding idiosyncratic avoidance of *septante*); *soixante* is higher-valued than both *cinquante* and *quarante*. A similar argument also works, or course, for *quatre-vingt-dix*, (4 × 20) + 10, 90, which is not expressed as **trois-vingt-trente*, (3 × 20) + 30, although general rules which generate the former would also generate the latter.

An example from Mixtec[2] shows the recursivity of the phrase structure rules. Note also the conformity with the Packing Strategy which ensures that the sisters of the recursive NUM(BER) nodes are the highest-valued the lexicon of the language will allow, below the arithmetic ceiling value of the immediately dominating node. In this structure it is not possible to transpose any two different PHRs, or the two Ms, without losing well-formedness.



In this example, category names have been abbreviated, and appropriate numerical values subscripted to some of them. The whole expression means 7999.

One more highly recursive example, from Hawaiian[3] makes a similar point.



As mentioned, there are rare exceptions to the Packing Strategy, which help to illustrate its working. The key phrase in the Packing Strategy is ‘‘the highest possible value’’. This superlative phrase guarantees that for any position in a numeral expression only one subconstituent is

[2] The Mixtec example is from Merrifield (1968); the phrase structure analysis is from Hurford (1975).
[3] The Hawaiian example is from Anonymous (1834); the phrase structure analysis is from Hurford (1975).

possible, and further predicts that for any given number, a language will have a single unique well-formed expression. The existence of numeral paraphrases such as *two thousand*, *one hundred* and *twenty one hundred* are counterexamples to this prediction. Only the former should be acceptable, as *thousand* is a higher-valued M than *hundred*. English is idiosyncratic in outlawing expressions like **twenty hundred* and **fifty hundred*, in accordance with the Packing Strategy, but not outlawing expressions like *twenty one hundred* and *fifty seven hundred*.

But notice some significant facts about such a counterexample, which give us a clue to the proper status of the Packing Strategy. **Twenty hundred and ninety nine* is not well-formed, whereas *two thousand and ninety nine* is well-formed. And if an English speaker is asked to count forward from *two thousand and ninety nine*, the next enumerated expression will not be *twenty one hundred*. In the context of a recited counting sequence, a problematic counterexample to the Packing Strategy disappears. In English, expressions like *fifteen hundred* and *twenty six hundred* are frequently used in reading aloud expressions written in 'Arabic' notation like *1500* and *2600*, in which case, perhaps, an easy parsing of the string of digits gives salience to the *-00*, prompting use of the word *hundred*.

On another matter if idiosyncratic detail, English *eleven* and *twelve* are not now transparently bi-morphemic, so we might expect English to use a base of 12, as the highest available DIGIT. What appears to have happened here is that originally the English expressions for 11 and 12 were bi-morphemic, with the form for 12 explicitly, and the form for 11 implicitly, using a base of 10. The other '-teen' expressions have retained their transparent bi-morphemic status, but subsequent phonological erosion and modification selectively targeted for forms for 11 and 12. Note that *thirteen* and *fifteen* have also undergone some historical changes. A similar story can be told for French *onze*, *douze*, *treize*, *quatorze*, *quinze* and *seize*, which are arguably not now bi-morphemic, although they retain clear traces of the corresponding DIGIT forms.

Finally, before taking up attempts to explain the Packing Strategy, a note on why numeral expressions should be regarded as syntactic constructions, rather than (sometimes long) words, perhaps compounds, formed by morphological rules. In some languages the syntactic status of numerals is very clear, in that non-numeral expressions can be intercalated into their interior. The Celtic languages provide good examples. An example from the Welsh Bible is *ddeng mlynedd ar hugain ac wyth cant*, literally 'ten years on twenty and eight hundred', translated as '830 years'.

2. A previous attempt to explain the Packing Strategy

In Hurford (1987) a chapter was devoted to arguing that the Packing Strategy can be explained by a gradual socio-historical process of standardization. Early in the history of a language, it was argued, there would have been a variety of numeral expressions with the same value. Certain numbers would have been used more frequently than others, because higher numbers are more versatile in forming a range of compound expressions. As a basic example, consider ways of expressing numbers between 11 and 20, by addition of two numbers between 1 and 10. 15, for example, can be expressed as 10 + 5, 9 + 6, or 8 + 7; 17 can be expressed as 10 + 7 or 9 + 8; and 19 can only be expressed as 10 + 9. Over time, language learners would therefore hear more examples of numerals formed on a base of 10, and would at some point conclude that adding to 10 is the 'standard' way of forming numerals in this range.

In Hurford (1987) the internal consistency of this putative explanation was tested with a computer simulation, in which simulated agents in a population were given random numbers to express to each other. At the beginning the simulation (analogous to the beginning of the history of the numeral system), speakers just used any arithmetically correct combination, using addition and multiplication, and subject to certain natural economy constraints (e.g. prefer a shorter expression to a longer one, *ceteris paribus*). The agents in this simulation learned from the practices of the other agents, and after many generations the agents formed internal grammars of numeral expressions which severely constrained the possible numeral expressions. In Hurford (1987) these simulations succeeded in achieving outcomes conforming to the Packing Strategy for decimal systems up to 100. It was left to later work to determine whether this type of explanation would scale up to the higher reaches of numeral systems, and make the correct predictions about numerals involving thousands, millions and billions.

In the intervening years the format of the earlier simulations had been more extensively explored under the general heading of Iterated Learning Models. The Iterated Learning Model reflects a general theory of cultural language evolution. It is not limited to numeral constructions; it has been invoked to explain aspects of language such as general semantic compositionality, irregularity of frequently-occurring forms, and Zipf's Laws. (See Kirby, 2001; Kirby and Hurford, 2002; Kirby et al., 2004 for example and discussion.) The central idea is that languages exist in two modes: as public behaviour, utterances 'out there', and as private mental representations, grammars 'in the head'. These two modes cause each other, in a diachronic cycle: public utterances conform to private rules, and children learn their rules from public utterances.

For the present publication, the earlier simulations of the cultural evolution of the Packing Strategy (at least up to 100) were replicated, explicitly within the Iterated Learning framework. Similar results, illustrated here, were obtained. Given below is a sample of the initial random unruly data produced by agents in an early stage of the simulation.

6×9	4×10	3×4	$5 \times 6 + 1$	4×6	$6 \times 10 + 6$	5×8
$6 \times 6 + 7$	$5 \times (9 + 4)$	8×8	$6 \times 6 + 10$	6×10	8×9	$9 \times 10 + 3$
$3 \times 9 + 7$	$7 + 4$	$7 + 5$	$9 + 2$	$8 + 7$	$8 \times (7 + 6)$	$6 \times (4 \times 4)$
$9 \times (10 + 2)$	5×9	$7 \times 7 + 2$	4×6	$3 \times 8 + 2$	$10 \times 10 + 1$	4×8

Although these are arithmetically acceptable formulae, clearly many of them are linguistically very bizarre. For example, no language expresses 65 as $5 \times (9 + 4)$, or 72 as 6×9, or 51 as $7 \times 7 + 2$. On the other hand, there are some linguistically reasonable examples here, such as $6 \times 10 + 6$, or $9 \times 10 + 3$. In this simulation, each agent in each generation was exposed to a limited number of examples, and, on the basis of the relative frequency of examples of certain clearly defined types, internalized a grammar, consistent with the data experienced, but usually generating fewer paraphrase expressions than the agents in the previous generation. Given below is the whole series of numerals from 11 to 100, as one agent, after a few generations, would have expressed them.

$10 + 1$	$10 + 2$	$10 + 3$	$10 + 4$	$10 + 5$	$10 + 6$	$10 + 7$	$10 + 8$	$10 + 9$	2×10
3×7	$2 \times 10 + 2$	$2 \times 10 + 3$	**3×8**	$2 \times 10 + 5$	$2 \times 10 + 6$	**3×9**	**4×7**	$2 \times 10 + 9$	3×10
$3 \times 10 + 1$	**4×8**	$3 \times 10 + 3$	$3 \times 10 + 4$	**5×7**	**4×9**	$3 \times 10 + 7$	$3 \times 10 + 8$	$3 \times 10 + 9$	4×10
$4 \times 10 + 1$	**6×7**	$4 \times 10 + 3$	$4 \times 10 + 4$	**5×9**	$4 \times 10 + 6$	$4 \times 10 + 7$	**6×8**	**7×7**	5×10
$5 \times 10 + 1$	$5 \times 10 + 2$	$5 \times 10 + 3$	**6×9**	$5 \times 10 + 5$	**7×8**	$5 \times 10 + 7$	$5 \times 10 + 8$	$5 \times 10 + 9$	6×10
$6 \times 10 + 1$	$6 \times 10 + 2$	**7×9**	**8×8**	$6 \times 10 + 5$	$6 \times 10 + 6$	$6 \times 10 + 7$	$6 \times 10 + 8$	$6 \times 10 + 9$	7×10
$7 \times 10 + 1$	**8×9**	$7 \times 10 + 3$	$7 \times 10 + 4$	$7 \times 10 + 5$	$7 \times 10 + 6$	$7 \times 10 + 7$	$7 \times 10 + 8$	$7 \times 10 + 9$	8×10
9×9	$8 \times 10 + 2$	$8 \times 10 + 3$	$8 \times 10 + 4$	$8 \times 10 + 5$	$8 \times 10 + 6$	$8 \times 10 + 7$	$8 \times 10 + 8$	$8 \times 10 + 9$	9×10
$9 \times 10 + 1$	$9 \times 10 + 2$	$9 \times 10 + 3$	$9 \times 10 + 4$	$9 \times 10 + 5$	$9 \times 10 + 6$	$9 \times 10 + 7$	$9 \times 10 + 8$	$9 \times 10 + 9$	10×10

While most of these expressions are typical of a regular decimal system, this agent has a preference for economical non-decimal expressions, highlighted above in boldface, such as

3 × 7, 7 × 8 and 9 × 9. After more generations of the iterated learning process, all such non-decimal multiplicational expressions were eliminated from the community's language. In 20 runs, under certain conditions, the simulations always converged on rules which generate English-like numerals up to 100, conforming to the Packing Strategy.

So far, so good, but in fact this type of explanation does not scale up. It has not been possible to push it to produce numerals systems reaching above 100, conforming to the Packing Strategy. The learned rules are purely local, like context-free phrase structure rules generally, in the sense that each rule specifies the syntactic types of the immediate daughters of a dominating node, with no reference to any properties of the sisters of the dominating node. The Packing Strategy, of course, determines the overall well-formedness of a numeral expression specifically in terms of arithmetical relations between sister constituents (''The sister of a NUMBER must have the highest possible value.'') The ''highest possible'' condition here also requires a search of other structures generated by the grammar, a computationally highly costly process. It is hard to see how, without allowing global searching rules of this sort into each agent's competence, a population of agents could be engineered to converge on an extended numeral system conforming to the Packing Strategy. The model will not scale up to complex constructions with more layers than the relatively simple expressions up to 100.

3. Counting practices explain the Packing Strategy

All previous discussion of the Packing Strategy has been conceived within the broad framework of generative grammar. From this perspective, the numerals in a language, just like the sentences, are an unordered set. But numerals are unique among linguistic constructions in being ordered. *Eight* only means 8 because it is the eighth item in a conventional recited sequence. Few other lexical items are like this. Another example is the sequence *Monday*, *Tuesday*, *Wednesday*, *Thursday*, *Friday*, *Saturday*, *Sunday*. A day is only called *Friday* if it follows a day that was called *Thursday*, and we only know this because we know the recited sequence of day names. But numeral expressions, uniquely, extend the significance of sequence beyond lexical items, into the compound expressions formed from them. Children are first taught the recited lexical sequence, and then taught how to continue this sequence using syntactically constructed expressions. In this exercise, the fact that they are learning a sequence is always implicit. And of course, the *raison d'être* of the whole sequence is its practical application in counting things. In this section, I will suggest that the Packing Strategy can be explained because numerals are specifically designed to suit the practical exercise of counting.

We need to put ourselves in a very practical frame of mind. Imagine a collection of objects on a surface in front of you. You have to count them. You know a recited sequence of counting words. You separate out each object in turn, moving it to one side, to show that it has been counted, and with every such move you utter the next word in your memorized sequence. You go as far as you can this way. If there are no more objects than you have counting words in your memorized sequence, all is well, and you have counted the objects. For example: ''One, two, three, four, five, SIX! Aha, there are six of them.''

The important idea here is expressed by ''Go as far as you can with the resources you have''. A person with a counting sequence that ends at 5 can only count collections of 5 or less by this method. Now, keeping in this same practical frame of mind, imagine that there are more items to be counted than you have words in your recited counting sequence. You come to the end of the sequence, and this time move a whole subgroup aside, gathered tightly together into what can now be regarded as a unit. You know how many items this unit contains, and the last word in your counting sequence can be used as a predicate to describe it. For concreteness, we will use English words, so this subgroup gets called a *ten*. Then you count the remainder, starting again at the beginning of your recited sequence. Let's say that now you get to *four* as you finish counting all the objects. You now have two subgroups, which are naturally called a *ten* and a *four*. Universally across languages, there is a very strong correspondence between conjunctions used to express groupings of groups, as in *these apples and those pears*, or *this team and that team*, and arithmetical addition. So it is natural to denote the pair of groups counted out as something like *a ten and a four*. Singular forms of the predicates are appropriate, as the properties of 'tenness' or 'fourness' belong to the groups, and not to the individual members (unlike apples). But in English at least, plural modifiers are still allowed, so that one can say either *that ten* or *those ten*.

Some objects come naturally in groups of regular cardinality, but there are very few such objects and the numbers involved are invariably low. Some bodyparts, such as eyes, ears, arms and legs, come in twos, as do the corresponding items of clothing, such as shoes or gloves. When counting people at a social gathering, it might sometimes be appropriate or convenient to count them by couples. But generally the world enforces no strong preference for any particular number of objects in a group. The groupings made during counting are usually ad hoc and for the convenience of the counter. If the items to be counted are not naturally grouped somehow, the counting practice imposes a grouping into subcollections that are the size of the recited counting sequence. The maxim ''Go as far as you can'' means that, given a recited sequence up to *ten*, it would be unnatural to count out a uniformly distributed group of 14 objects as two groups of seven, or as a nine and a five, or an eight and a six. We can see the beginnings of the Packing Strategy here.

The practical maxim ''Go as far as you can with the resources you have'' is seen in contexts outside counting. When picking apples in an orchard and carrying them home in baskets, it is natural, and efficient, to fill each basket to the top, or to the maximum carryable weight. Another possible example involves walking in a city built on a grid plan. If you need to get to a place several blocks east and several blocks north of your present position, the available shortest routes include many zigzag routes alternating eastward and northward movement, turning at many corners. But the conceptually simplest route involves going straight in one direction as far as necessary, and turning only once. Certainly, in giving verbal directions, the simplest advice would be of this ''Go as far as you need'' type. You pack all the movement in one direction into a single continuous stretch, before embarking on movement in the next direction.

I have mentioned the natural correlation between a conjunction used to signal a grouping of groups and a conjunction used to signal addition in numeral constructions, as with English *and*. Another natural correlation is between pluralization and multiplication. Note that while *that ten* and *those ten* are close paraphrases, *those tens* means something different; in this case we are talking about more than one group of ten objects. Generally across languages the numeral words used as multiplicative bases, such as *hundred*, *thousand* and *million*, are nouns, and are often pluralized, as in French *quatre millions*. With the lower-valued numeral bases, there can be exceptions to this generalization, but statistically across languages, there is a correlation between high value of a multiplicative base and its nouniness in the grammar.

Now returning to the example of practical counting, say we have a quite large collection to count. We group the objects into tens, leaving any remainder, and we count the tens using our recited lexical sequence. In this way we might arrive at *four tens and six* or *nine tens and nine*, for example, or even at *ten tens*. At this point a shift happens. In developed decimal numeral systems, a new lexical item is introduced with the meaning *ten tens*. In English, this is *hundred*. Now imagine counting an even larger collection of objects. We put groups of ten aside, as before, and when we have ten of them, we form a larger group and call it a *hundred*. We might put several hundreds aside in this way, until we are left with a remainder with less than a hundred objects. In this way, using conjunction and pluralization, we could arrive at expressions such as *six hundreds, four tens and three*. It is easy to see, given a new word for *ten hundreds*, how such an example can be continued to give expressions such a *five thousands, two hundreds, four tens and six*. The expressions arrived at in these counting examples all conform to the Packing Strategy.

The structure for English *three hundred thousand* implies a count of 300 groups of 1000. How does a normal counting procedure arrive at this, rather than 3000 groups of 100, corresponding to the ungrammatical **three thousand hundred*? We will work through the example. You count the first 100 objects, putting them aside into a group labelled *hundred*. Carry on thus until you have put aside ten groups labelled *hundred*. Now here there is apparently a choice between simply carrying on accumulating more groups of 100, or forming a supergroup out of the ten hundreds and calling it *thousand*. I suggest that another natural principle applies here, namely ''Minimize the number of entities you are dealing with''. Amassing the ten collected hundreds into a single larger group cuts down on the number of entities being dealt with. The grouping process has the effect of shifting individual objects out of cognitive focus, and replacing them with a single cognitive unit. So at this apparent choice point, the ''Minimize entities'' principle dictates that you form a supergroup out of the ten collected hundreds and call it a *thousand*. After that, following the same principles, you will eventually arrive at 300 groups of 1000, and not at 3000 groups of 100.

Very few people in fact ever have to count such large collections of objects. Indeed, perhaps no-one has ever verbally counted as many as 300,000 evenly distributed objects. The two principles I have proposed, ''Go as far as you can'' and ''Minimize entities'', have to be seen as ideal conceptual guidelines, which are indeed followed for counting up to smallish numbers. Users, I claim, become aware of applicability of these principles to counting in general, up to very high numbers.

The counting practice described above is structurally similar to a well-known, and very simple sentence parsing algorithm, known as ''Shift-and-Reduce''. In this algorithm, the parser starts at the beginning of the sentence, takes in the first word, and ''reduces'' it by looking up its lexical category. It then ''shifts'' to the next word, and reduces that. Having now got two lexical categories, it may be in a position, depending on the grammar of a language, to further reduce these two lexical categories to a single phrasal category. In parsing *The cat sat on the mat*, for example, first *the* is reduced to DET, then *cat* is reduced to N, and DET N is reduced to NP. The counting process I have described goes through a collection of objects, starting at some arbitrary object, and having accumulated a group of a certain size, reduces this to a numeral term, e.g. *ten*, then carries on shifting through the collection until another reducible subgroup (e.g. another *ten*) has been put aside. At a later point, if the collection is large enough, ten of these *tens* will be reduced to *hundred*. It can be easily seen how the process would continue for very large numbers. In this analogy, the higher-valued multiplicative bases are analogous to phrasal categories constructed by the parser. In a broad general sense, counting objects is like parsing a part of the world, reducing it to linguistic order.

The Packing Strategy depends for its operation on the lexical resources provided by the language. In our example so far, we have imagined a regular decimal system, with lexical items for the values 100 and 1000. Languages sometimes introduce lexical items earlier than they are strictly needed. For example, Welsh has a word *pymtheg*, 15, with which expressions between 16 and 19 are formed: *un ar bymtheg*, 1 on 15, *dau ar bymtheg*, 2 on 15, *pedwar ar bymtheg*, 4 on 15. Given the availability of this word in the language, it is natural to use it in the counting sequence as a base to which other smaller numbers can be added. But the usefulness of *pymtheg* is swiftly curtailed by the next lexical item provided by the language, namely *ugain*, 20. After 20, Welsh counts in twenties, not in fifteens, up to 100, when *cant*, 100, becomes available.

The Hawaiian example given earlier showed that this language has lexical items for 400,000, 40,000, 4000, 400, 40, 20 and 10. Counting in Hawaiian up to very large numbers, the principles ''Go as far as you can using the resources you have'' and ''Minimize entities'', applied at each stage, would yield the high-valued numeral expression of that example, conforming to the Packing Strategy.

There is no contradiction between the ''Go as far as you can'' principle and the fact that languages sometimes introduce lexical items earlier than they are strictly needed. Counting is done by individual language users, using the resources their language makes available. The coining of new lexical items is a much more sporadic process. In some cases, contact between languages provides a language with a lexical item which it did not have before, and which would not previously have made a natural grouping during the counting procedure. This has frequently happened when a vigesimal system is invaded by a decimal system, and a word for 100 becomes available.

I have given the outcomes of hypothetical counting examples in stylized forms such as *two hundreds, three tens and four* or *three hundreds of thousands*. The numerals of some languages are as clear and transparent as this. But other languages in the course of their history have modified the expressions built on such a framework. Thus we find: morphology taking over from syntax, as in English *twenty*, *thirty*, . . ., *ninety*; differential pluralization of different multiplicative bases; truncation, as in Danish *halvtreds*, contracted from *halvtredsindstyve*, 'half third times twenty' (50); and many other idiosyncratic processes. None of this relates to the argument of this paper, which is about the central arithmetical scaffolding of numeral expressions.

4. In conclusion

Perhaps this conclusion was obvious, that a universal property of compound numeral expressions derives from the main practical task that numerals serve. The task is very clear and well known. Many functionalist arguments exist that certain linguistic constructions derive their specific properties from the complex meanings they typically express. But the closeness of detailed correspondence between the task and the linguistic patterns is not usually, or perhaps ever (yet!), as great as the correspondence between the practical counting task and the Packing Strategy. Numeral systems are far from typical of the rest of language. They are in some sense a new technology. Children are explicitly taught to count, whereas they acquire most of the rest of their language without explicit instruction.

I have often wondered whether analogues of the Packing Strategy exist outside numeral systems. This now seems unlikely, as the practical task from which the strategy derives is so special.

References

Anonymous French Catholic Missionary, 1834. Notes Grammaticales sur la Langue Sandwichoise. British Museum catalogue no. 12901.bb.47, Paris.

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

Hurford, James R., 1987. Language and Number: The Emergence of a Cognitive System. Blackwells, Oxford.

Kirby, S., 2001. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation 5 (2), 102–110.

Kirby, S., Hurford, J., 2002. The emergence of linguistic structure: an overview of the iterated learning model. In: Cangelosi, A., Parisi, D. (Eds.), Simulating the Evolution of Language. Springer Verlag, London, pp. 121–148.

Kirby, S., Smith, K., Brighton, H., 2004. From UG to Universals: linguistic adaptation through iterated learning. Studies in Language 28 (3).

Merrifield, W.R., 1968. Number names in four languages of Mexico. In: Corstius, H.B. (Ed.), Grammars for Number Names. Foundations of Language, Supplementary Series, vol. 7. D.Reidel, Dordrecht, Holland.