More than words

**Geoffrey K. Pullum and
Barbara C. Scholz**

In the popular view, a language is merely a fixed stock of words. Purists worry about foreign loanwords; conservatives decry slang; and groundless claims that there are hundreds of Eskimo words for snow are constantly made in popular writing, as if nothing matters about languages but their lexicons.

But the popular view cannot be right, because (as linguist Paul Postal has observed) membership in the word stock of a natural language is open. Consider this example: "GM's new Zabundra makes even the massive Ford Expedition look economical." If English had an antecedently given set of words, then this expression would not be an English sentence at all, because 'Zabundra' is not a word (we just invented it). Yet the sentence is not just grammatical English, it is readily interpretable (it clearly implies that the Zabundra is a large, fuel-hungry sports utility vehicle produced by General Motors). Similar points could be made regarding word borrowing, personal names, scientific nomenclature, onomatopoeisis, acronyms, loaned words, and so on; English is not a fixed set of words.

A more fundamental reason that a language cannot just be a word stock is that expressions have syntactic structure. For example, in most languages, the order of words can be significant: "Mohammed will come to the mountain" contains the same words as "The mountain will come to Mohammed", but the expressions are very different. Inclusion within phrases is also an important part of syntactic structure. In the expression "We could not tell him", ambiguity arises from the fact that 'not' may belong in the same phrase as 'tell him' — in which case the meaning is that keeping him in the dark is possible or permitted — or it may be outside the 'tell him' phrase, in which case 'not' belongs with 'could' and the meaning is that telling him is impossible or forbidden.

The syntactic structure of natural languages has several important features. One is revealed by the example just considered; there is no guarantee of what mathematical logicians call 'unique readability' — there is no one-to-one correspondence between sound strings and syntactic structures, or between syntactic structures and meanings. Natural languages are replete with ambiguity.

A second feature is that there is no upper limit on the complexity of expressions. A verb phrase such as 'run away' can be embedded in a larger verb phrase such as 'see Spot run away', and there is no syntactic limit on further embedding, so expressions can be of arbitrary complexity: "Tell him they think he overheard someone ask her to confirm that they saw him watching us waiting for you to see Spot run away in order to …". Hence, natural languages are productive, as they possess the structural resources for indefinite recombination.

A third feature of natural language syntax is its variability. Even within a single speech community (which we can roughly define as a human group whose members broadly understand each other's speech and recognize it as being characteristic of the group), there are quite sharp differences concerning the relevant regularities of syntactic structure, both between subgroups (dialect differences) and between individuals, whose idiosyncratic divergences mostly go unnoticed.

Fourth, malformations of syntax vary in their severity — some partially ill-structured expressions are more deviant than others. President George W. Bush's departures from standard English syntax are well known; Tarzan's departures are more extreme, and Yoda's even more so ("Already know you that which you need"); yet the structure of English is partially respected in each case. Likewise, familiar phenomena such as hesitation ("It's in the … in the drawer") and use of fragments ("And that would be …?") partially conform with English expressions; they are not random jumbles of words.

VICKY ASKEW

Language

Natural languages are unique in combining the best of formal languages and primate communication.

Within mathematical logic and computer science, invented formal symbolic systems are called languages, but in some respects they are strikingly different. Their syntactic structure allows embedding and re-embedding, but their vocabularies are fixed; ambiguity is ruthlessly excluded; structures must be completely well-formed, and disrupted or fragmented expressions simply do not belong at all. Most of the linguistic work on syntactic theory over the past 50 years has used the mathematical methods devised for defining formal languages of this sort. This suggests a possible confusion of formal tools with subject matter.

To formulate grammars that describe natural languages precisely, we need a formal metalanguage that has the properties of any other scientific, formal language: a recursively defined, unambiguous syntax that defines a countably infinite set of expressions. But natural languages themselves — the focus of scientific linguistics — are not precisely delineated sets of expressions, any more than they are precisely delineated sets of words.

The aspects of language that can be learned by certain great apes, particularly the chimpanzees *Pan troglodytes* and *Pan paniscus*, are curiously the very same ones that fit the popular conception of languages. Apes can learn to name things using hand signs or visual symbols, and to express some basic demands by uttering them ("Open fridge! Give apple give!"), but they seem to be incapable of developing a productive grasp of syntax.

Human languages exhibit a unique combination of characteristics: first, semantic word-to-world relations that we share with other primates; second, syntactic structures as complex and exact as in formal languages; and third, an openness, flexibility and ambiguity that formal languages do not allow. ∎

Geoffrey K. Pullum is in the Department of Linguistics, University of California, Santa Cruz, California 95064, USA. Barbara C. Scholz is in the Department of Philosophy, San José State University, San José, California 95192, USA.

FURTHER READING

Karmiloff, K. & Karmiloff-Smith, A. *Pathways to Language* (Harvard Univ. Press, Cambridge, MA, 2001).

Pullum, G. K. & Scholtz, B. C. in *Logical Aspects of Computational Linguistics, 4th Int. Conf.* (eds de Groote, P., Morrill, G. & Retoré, C.) 17–43 (Springer, Berlin, 2001).

Wallman, J. *Aping Language* (Cambridge Univ. Press, Cambridge, 1992).