

D. Terence Langendoen and Paul M. Postal: *The Vastness of Natural Languages*, Oxford, Basil Blackwell, 1984, 189 pp.

The book under discussion does not open a new era in the development of linguistics, but it certainly opens a new area of discussions on the character of natural language both from a philosophical and a mathematical viewpoints. The main conclusions of the authors consist in:

(i) The sentences of natural language exhibit no size limitation, i.e. natural language includes not only infinite, but also non-denumerably long sentences, since coordinate constructions of the shape *Babar is happy and I know that Babar is happy and I know that I know that Babar is happy and...* without any size limits meet all the structural conditions of sentences (the language is closed under coordinate compounding, see esp. pp. 53-59), and Occam's razor excludes any additional bounds.

(ii) From (i) it follows that natural language, understood as a class (collection) of sentences, is no set, i.e. it is a proper class, or a megacollection (p. 8f); it is bigger than countably infinite (esp. 65f).

(iii) From (ii) it follows that natural language has no constructive (generative, Turing-machine-like) grammar (p. 72).

(iv) Chomskyan descriptions, based after 1980 on "radical conceptualism", assume that grammar generates mental representations of sentences (thus differing from "standard conceptualism", which suggests that sentences themselves are generated); this is untenable, since the existence of grammars that have not yet been learned by any individual would then be excluded (pp. 125ff).

(v) Since also other conceptualist approaches to natural language meet serious difficulties, Katz's understanding of

natural languages as platonic objects is to be preferred; linguistic objects can then be studied as abstract real entities, i.e. the core of linguistics is a logico-mathematical discipline (p. 159n).

(vi) The only nonconstructive type of linguistic description elaborated up to now is Arc Pair Grammar (formulated by Johnson and Postal); this framework can be modified in such a way as to correspond to the absence of any limitation (finite or transfinite) on the length of natural language sentences.

The book under review has already been discussed thoroughly by S. G. Lapointe (1986). One of the main objections raised there consists in the fact that, if the empirical aspect of linguistic theory (substantially influenced by the requirement of adequacy with respect to the observed facts) is paid due attention, then the arguments based on Occam's razor (as basic for the simplicity of theoretical statements) are not the only relevant considerations. In deciding between competing theories, not only the relative simplicity of the theory should be decisive, but also the requirement that such a theory should be chosen that places the greatest restrictions on the structure of the empirical domains (this concerns those consequences of the theory which no known facts at present determine as being true or not). A constructive system, although it is not the only possibility how to constructively describe the set of attestable sentences, still is preferable, since it imposes "far more restrictions on what can count as a possible grammar for a natural language" than a nonconstructive system of the kind envisioned by Langendoen and Postal (see Lapointe, p. 232); this leads to the conclusion that the closure under coordinate compounding has a bound, not allowing for non-denumerably long sentences.

It should be added that a similar bound would have to be required also for other operations¹, since e.g. also the possibility to add adjectives (or relative sentences) to nouns in natural language sentences seems to undergo no structural limits. Thus the dispute between platonist and conceptualist view-

points appears not to be closed.

Lapointe certainly is right in that Occam's razor by itself represents no absolute criterion for the choice between alternative theories in science. If its application at one point of the reasoning leads to relatively complicated situations at other points, then it may be suitable to reconsider the given way of thinking altogether. If the use of Occam's razor to base every size limitation from the sentence leads to an understanding of natural language as a megacollection, then it might be useful to pay more attention to Keenan's remark (quoted on p. 45 of the book under review, cf. p. 227 in Lapointe's review) according to which the nature of constructive systems could be understood to be responsible for the finiteness of all natural language sentences. The requirement to elaborate this way of thinking more systematically and to give it more weight seems to be corroborated by the essential role of the human factor in all what concerns human language; e.g. one of the basic conditions might consist in that every sentence of a natural language be such that it can be produced by a human speaker; cf. also Chomsky's (1986, 55) "feasibility". In addition, one should consider also following points:

(i) Langendoen and Postal remark (p. 47, 141) that also morphs have no length bound, finite or transfinite; however, if structural properties are connected with any length bounds neither in the syntax, nor in the lexicon, can we be certain that there are bounds restricting the number of other linguistic units, be their way of combination based on an ordering or not? On what grounds can it be stated that every word form has just a finite or denumerable number of morphemes (cf. the free formation of compound nouns in English, German, or Sanskrit, and also the free addition of agglutinative morphemes e.g. in Hungarian)? What makes it certain that a single form has just a finite or denumerable number of functions, or, to speak in more general terms, that an item of a level has only a finite number of corresponding items on the adjacent level(s)?

(ii) Formal languages of logic usually are defined in a constructive way; however a similar argument (the absence

of structural differences between formulas of different length) applies there, too. Why do such questions attract so few attention in logic and mathematics, the objects of whose study certainly lack limitations based on "performance reasoning" on stronger grounds than those of empirical cognitive studies?

(iii) If the degrees of psychological realism of grammars are discussed, then a system consisting of (a) a non-constructive grammar, (b) a constructive parser, and (c) a constructive producer (cf. pp. 116ff in the book under review) might meet with certain difficulties; can we hope that the result of acquiring a second language will be described just by adding a new (non-constructive) grammar to the general procedures of production and of parsing which have already been acquired by the subject with her/his native language (or are innate)? However, the experience of the research in parsers (oriented towards automatic translation, or towards different models of natural language understanding) corroborates the view that such general (language independent) procedures would be rather ineffective, so that the efforts aiming at a simpler model (esp. a reversible grammar in the sense of Martin Kay) have not yet been abandoned. Or do we have to accept that a learner of a second language internalizes three more devices?

Even if it were firmly established that grammars are non-constructive, the statement that Arc Pair Grammar (with the relevant modifications) is the only extant non-constructive framework of linguistic description, is not so important as it might seem. A more basic question then is how difficult it is for other approaches to find a non-constructive formulation for variants of their frameworks; this concerns Chomsky's government-binding theory as well as Bresnan's lexical functional grammar, dependency (valency) based approaches with complex node labels, or other systems.

Another question that has to be - and is being - reconsidered concerns the meaning of the term 'language'. Chomsky's

use of 'language' for the set (or, should we now say, class) of sentences never has been quite generally accepted. The Saussurean identification of language (*la langue*) with a system (be it of restrictions, rules, principles, or "phenomena") never has been fully abandoned (cf. e.g. McCawley, 1976); it may be considered not to be crucial whether 'language' is used as a synonym of 'linguistic competence' or for the class of all sentences. What is essential, is to be aware of the difference, i.e. of the fact that here, as so often, science can differentiate between several explicata corresponding to different layers or aspects of a single explicandum. Chomsky (1986, p. 29; 150f) speaks now about such a difference. Perhaps a further differentiation can be achieved, e.g. by distinguishing between language₁ as linguistic competence, language₂ as the set of sentences of finite length, and language₃ as the megacollection of sentences of all lengths (the object of mathematical linguistics in the sense of Lapointe, 1986, 237).

The notion of language₂ may still be considered arbitrary in the given context; however, if it is taken into account that natural language is an attribute of the human species with its limitations, then it might be understood as basically important that the difference between "too large" and "infinite" ceases to be substantial in some recent trends in set theory, cf. esp. Vopěnka's (1979) approach to unsharpness, based on the fundamental role of the human horizon; the proper classes exhibit other interesting properties under such an approach than just their largeness.

We would like to add a remark that is not crucial for the given dispute but may be important in other respects. The word order instances illustrated by such examples as G. *Die Mutter sieht die Tochter* (see p. 119 in the book under review) probably are not adequately characterized by Chomsky, and a 'corpus rule' in the sense of Langendoen and Postal seems not to be sufficient here: After *Wer sieht die Mutter?* can *Mutter* well be understood as the object, similarly as *die Flasche* in *Die Flasche sieht die Tochter* after *Wer sieht die Flasche?*

As we have remarked, the questions posed by Langendoen and Postal in their book will certainly become the object of important discussions, which will concern not only the length of sentences, the constructiveness of the description, or the explication of the term 'language', but also the problem of the non-existence of a class of all human languages (mentioned by the authors on p. 86f), which has important implications for the typology of languages and for the theory of linguistic universals.

Petr Sgall

Note:

1 The claim that natural language cannot be understood as a recursively enumerable set of sentences was formulated already by Savický (1967).

References:

- Chomsky N. (1986), *Knowledge of Language, Its Nature, Origin and Use*, New York.
- Lapointe S. G. (1986), review of Langendoen and Postal, *The Vastness of Natural Language, Linguistics and Philosophy* 9, 225-243
- McCawley J. D. (1976), *Some Ideas Not to Live by, Die neueren Sprachen* 75, 151-165
- Savický N. (1967), O možnostech matematického modelování sémantiky přirozeného jazyka (On Possibilities of Mathematical Modelling of Natural Language Semantics), *Kybernetika* 3, 90-93
- Vopěnka P. (1979), *Mathematics in the Alternative Set Theory*, Leipzig (Russ. transl.: *Matematika v alternativnoj teorii množstv*, Moscow 1983)