

Analogical Modeling of Language

Royal Skousen

(Brigham Young University)

Dordrecht: Kluwer Academic
Publishers, 1989,
xii + 212 pp. Hardbound, ISBN
0-7923-0517-5

Reviewed by

Mirosław Bańko

University of Warsaw

This book addresses the general question of whether language is governed by rules or by analogy. Skousen's contribution to the long-standing debate on the subject is of primary importance, for he brings new arguments and a new kind of evidence into the discussion. Broadly speaking, Skousen constructs a computational model of a certain kind of analogical inference and tests it against empirical data in a series of experiments. This allows him to demonstrate that analogy accounts for what has been traditionally described by means of rules, and furthermore, that it predicts instances of variation and historical change that are otherwise difficult to explain.

In order to appreciate appropriately the import of Skousen's book, an historical background is needed. Unfortunately, Skousen draws a rather simplistic picture of the rules-versus-analogy controversy, paying more attention to conceptual distinctions than to historical facts. Generally, he contrasts his analogical approach to language with a broadly understood structuralist approach, which he attributes to de Saussure, Chomsky, and the Young Grammarians. However, while Chomsky indeed conceives of language as a rule-governed device and discredits analogy as a vague metaphor (Chomsky 1966, p. 12), de Saussure was well aware of the role of analogy in language use, language change, and language acquisition. Among the Young Grammarians, Paul too recognized the functioning of analogy in language (cf. Jespersen 1922, pp. 44–45), and so did many other linguists in the past (see Anttila 1977 for a state-of-the-art report).

Obviously, Skousen's approach to analogy is radically different from that of the older writers. One difference is that he gives an explicit definition of analogy, as if in response to a criticism such as Chomsky's (cf. above) and Langacker's (1967, p. 22), who remarks that since both grammatical and ungrammatical sentences may be the outcome of analogical inference, analogy is not a credible explanation of the human ability to create and understand novel sentences. A weak point in Langacker's argument is that he derives his "false analogies" from individual sentences, e.g. *I like kittens, I like snow, Kittens are white*, hence **Snow are white*. In Skousen's model, on the other hand, analogical inference is based on all relevant examples of usage stored in the speaker's mind.

Another difference between Skousen's approach to analogy and the traditional one results from his attitude toward rules. These are not used in his model, in contrast to the usual practice of employing rules to account for what is regular in language and invoking analogy only when irregularities have to be explained. A more recent view, that the rules/analogy distinction corresponds to the competence/performance dichotomy (e.g. Smith and Wilson 1979, pp. 202–203), is not shared by Skousen either. He makes his position quite clear in the last chapter, where he claims that rules are not

used in human language processing, and owe their existence entirely to linguists who invent them. He gives no reason, however, for this strong claim, except for a certain analogy he draws between linguistics and physics.

Perhaps the most serious drawback of Skousen's book is that it does not provide a sufficiently general description of his analogical model. Evidently, Skousen expects his readers to learn from examples how the model works and why it has been designed the way it has. True, he gives some details of the statistical basis of his approach, but these are very sketchy, probably for reasons of space. A Pascal program written for one of the applications of the model is by no means a substitute for a general, preferably algorithmic, description. Similarly, a reference to another book by Skousen, *Analogy and Structure*, does not help much, since the book has not been published yet.

Worst of all, it is unclear for a rather long time which particular kind of analogy is modeled in *Analogical Modeling of Language*. Again, only examples show that predictive analogy is concerned. The task of Skousen's model, given a context specified as a list of variables, is to predict the language behavior of the speaker (the outcome). The task is performed in two steps. First, an analogical set is generated that consists of all context–outcome pairs stored in the speaker's memory and relevant to the given context. Then a rule of usage is employed in order to predict the outcome for the given context on the basis of the analogical set. For instance, social variables such as the sex, age, and social class of the speaker are used in one of the experiments, and the outcomes predicted are forms of address in colloquial Egyptian Arabic. In another experiment, the English indefinite article is predicted, and the contexts are specified by listing particular phonemes and their category membership.

There are several limitations of Skousen's model, not all of them mentioned in the book. First, the model has been shown to cope well with predicting the speaker's language behavior, but it is less clear whether it could also account for the hearer's ability to understand novel sentences. One possibility is to put sentences into the contexts and to treat their semantic representations as the outcomes. However, none of the experiments described in the book is concerned with modeling natural language understanding, and the question of whether Skousen's model could be used to this end is left open.

Second, the list of variables used to specify the contexts has to be defined in advance, and it is the experimenter who decides which variables should be included in the list and which may be not. True, the exact role of particular variables need not be known initially, and in principle there is nothing wrong in considering variables that will eventually prove irrelevant to the speaker's behavior. No relevant variable, however, may be omitted from the list, which means that though independent of rules, Skousen's model is still dependent on some general knowledge of a given language, albeit expressed in a less direct form. It is an interesting question of how the model could be developed in order to acquire the knowledge itself.

Third, the time and memory required by Skousen's algorithm for generating the analogical set grow exponentially with the number of variables. The running time can be reduced by using parallel processing, but it was memory rather than time limitations that forced Skousen, working with a 640 KB computer, to use not more than twelve variables in any of his experiments. This small number proved to be big enough to test Skousen's model on the data usually considered by language variationists, but it may be insufficient for more complex tasks, such as predicting the form of a verb required by a given coordinated subject.

Despite its shortcomings, the book under review merits the attention of everyone having some theoretical interest in natural language modeling. Bold and novel, Skousen's analogical approach is a worthwhile alternative to the connectionist models

of language that have developed over the last ten years (e.g. Rumelhart, McClelland, and the PDP Research Group 1986). Of course, further research on analogical inference and its computer implementation is necessary. The limitations of Skousen's model mentioned above indicate the direction of this research.

References

- Anttila, Raimo (1977). *Analogy*. The Hague: Mouton (Trends in linguistics; State-of-the-art reports 10).
- Chomsky, Noam (1966). *Cartesian linguistics: A chapter in the history of rational thought*. New York: Harper & Row.
- Jespersen, Otto (1922). *Language: Its nature, development and origin*. London: George Allen & Unwin Ltd.
- Langacker, Ronald W. (1967). *Language and its structure: Some fundamental linguistic concepts* (second edition). New York: Harcourt Brace Jovanovich, Inc.
- Rumelhart, David E.; McClelland, James L.; and the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: The MIT Press.
- Saussure, Ferdinand de (1966). *Course in general linguistics*. New York: McGraw-Hill. (Original French edition 1916.)
- Smith, Neil and Wilson, Deirdre (1979). *Modern linguistics: The results of Chomsky's revolution*. Harmondsworth: Penguin.

Mirosław Bańko, who is a lecturer at the Institute of Polish Language, University of Warsaw, has a Ph.D. in linguistics and an M.Sc. in computer science. His recent research has concentrated on the algorithms for morphological synthesis and analysis, as well as on natural language learning systems. Bańko's address is: Instytut Języka Polskiego, Uniwersytet Warszawski, Krakowskie Przedmieście 26/28, 00-325 Warszawa, Poland; e-mail: banko@plearn.bitnet