From Grammar to Science: New Foundations for General Linguistics

Victor H. Yngve
(The University of Chicago)

Amsterdam: John Benjamins Publishing
Company, 1996, xi+350 pp; hardbound,
ISBN 1-55619-515-X, $89.00

Reviewed by
Geoffrey Sampson
University of Sussex

In the course of his long career, Victor Yngve has become seriously disenchanted with the unempirical character of modern linguistics. Within the discipline as it has developed,

> there seems to be no scientific way of deciding among the many
> contenders.... We find positions and methods being promoted like
> a new movie or defended with withering polemics or taken up like
> the latest fad ...

Most readers will recognize what Yngve is complaining about. For many of us who see it as a serious problem, the response has been to become "corpus linguists." The implications of the phrase seem a little odd (surely we would not expect to find that geography, say, included a specialism of "data geography" alongside a larger band of geographers who based their theories on guesswork or "intuition"?); but in practice "corpus linguistics" functions as a convenient shelter beneath which scientific linguistic research can proceed, undisturbed by the weird and wonderful things happening in areas of linguistics where people make their examples up out of their heads.

To Yngve, this is an insufficiently radical response to the empiricism problem. Even corpus linguists routinely depend on concepts such as "word," "noun," "sentence," and "utterance," which have no clear correlates in observable reality. Yngve surveys the origins of grammatical thought in the classical tradition and points out, correctly, that the grammatical categories still used today derive largely from concepts developed in connection with universal, a priori laws of logic rather than with description of contingent, concrete realities. He therefore sees any discourse that uses these concepts as doomed to be unscientific. Yngve's solution is that we should "abandon ... logical-domain theories entirely and mov[e] to the physical domain." We can observe the physical patterns of airwaves transmitted from speakers to hearers, as well as observing people as mobile bodies and the inanimate furniture of the environments in which speech occurs. Yngve outlines an abstract notation suitable to record lawlike relationships among properties of these physical complexes. "Because this notation can be programmed on a computer it can be used to test large-scale ... models"; and, because it refers to fundamental phenomena below the level of linguistic controversies, it "can be freely shared among different linguists.... Gone will be the babel of arbitrary grammatical notations, each to be discarded in turn."

Yngve deserves credit for taking seriously the negative implications of these ideas for his own past work. By far Yngve's best-known contribution is his Depth Hypothesis (Yngve 1960, 1961), which identified an asymmetry in the incidence of left-branching and right-branching grammatical structures. In the present book, Yngve discusses this

hypothesis at length, but concludes that "the depth hypothesis, as originally formulated [and he gives no new formulation], cannot be tested scientifically. Science rejects untestable hypotheses."

But the fact that linguistic terminology originated in logical discourse does nothing in itself to establish the unscientific status of modern linguistics. Quite a lot of sound science can trace its roots to abstract philosophizing in an earlier age. Four centuries ago, people "knew" that the planets travelled in circles, since this was the only geometry fit for inhabitants of the heavenly realm; that does not turn twentieth-century astronomers into unscientific phantasists. Yngve's Depth Hypothesis was highly testable in the Popperian sense that it identified potential phenomena which, if observed, would refute it. Indeed (although Yngve does not mention this), it has periodically been tested and refined. Fodor, Bever, and Garrett (1974, pp. 414–415) claimed, contrary to Yngve, that languages differ in their degrees of left-branchingness, giving measures to several decimal places for four languages. (Regrettably, there is a gap in their literature citation at this point.) Using corpus data of a kind not available to Yngve in 1960, I recently found (Sampson 1997) that the precise constraint in English is subtly different from what Yngve envisaged, and statistically exact in a way that I could not have anticipated before doing the research. If this was not an empirical scientific finding, I do not know what type of discourse it is—it is nothing like a priori logical analysis.

Yngve's critique of current linguistics might seem compelling if he made a good case for the superior empirical status of his "physical-domain" alternative. But we are told little about the specific properties that would be filled in to turn Yngve's abstract notation into a substantial science. Quite a lot of the discussion involves examples not relating to language, such as the binary property of being or not being "it" in the children's chasing game that Yngve calls "tag" and I call "touch." Yngve says that a hearer's property of being "on the spot to answer a question" (that is, of recognizing that one has been asked a question and the conversational ball is in one's own court) is similarly binary:

> We might be tempted to say that Sue's change from being not on the spot to being on the spot is a gradual continuous change starting when Tom says "Why ... " and ending when he finishes asking the question. But treating it as a continuous variable does not make sense because we cannot see anything but a binary distinction in [Sue]'s change in behavior.

Yngve seems to imply that being "it" in a game of touch or being "on the spot to answer a question" are theory-free predicates directly tied to observable correlates in a way that terminology such as "utterance" or "noun" is not. That is surely untenable. In the question case, we are not told what question Tom asks, but it is very easy to imagine cases where Sue's behavior changes observably before Tom reaches the end of his question (or where she ignores the question when it is complete).

Where Yngve's examples are linguistic, they are often very untypical. Many pages are devoted to the use of language between waitresses and cooks in a fast-food outlet where only 18 valid messages are available, but Noam Chomsky rightly pointed out 40 years ago that it is of the essence of natural language grammar to permit more than a small finite number of combinations. Most damagingly of all, to my mind, while Yngve stresses that speech is a "physical-domain" airwave phenomenon, in practice he takes for granted that it can be represented by words transcribed in the ordinary way. Anyone who keeps abreast of research on automatic speech recogni-

tion knows that the relationship between physical speech signals and orthographic forms is massively complex—even though researchers in that domain do assume that, much of the time at least, speech signals are intended as realizations of specific sequences of words, which for Yngve is an impermissible assumption. The community of speech researchers are investigating a very difficult domain in a thoroughly scientific spirit, without the faddishness or personality cults of which Yngve understandably complains in connection with theoretical linguistics. It is anyone's guess what speech researchers could achieve, if they were required to relate airwave patterns not to dictionary words but to observable changes in hearers' behavior; my guess is that they could achieve virtually nothing.

Yngve's basic error lies in moving from the proposition that linguistics needs to be more empirical to the proposition that it needs to be methodologically more or less identical to physics. Sciences of different domains are, and must be, different. Yngve explicitly asserts that the paradigm for any empirical science should be "the so-called hard sciences of physics, chemistry, and much of biology." I do not know where he means to draw the line between paradigm-worthy and other aspects of biology, but one very central area of that science is evolutionary biology, and it is notorious that evolutionary biology fails to conform to models of scientific method suggested by other hard sciences. The evolution of species resembles the history of a nation, as a succession of unique events, more than a physical system in which conjunctions of the same variables recur again and again; and the fundamental principle of survival of the fittest is arguably circular, since fitness may not be definable independently of survival. (These matters are discussed well by Ruse [1973], for example.) Do we conclude that evolutionary biologists are just fooling around rather than behaving as responsible scientists? Surely not; they are conforming as well as possible to the precepts of empirical science, in a domain in which that is not as easily done as in physics or chemistry.

In the linguistics of the last third of the twentieth century there has been a genuine problem, but it is not as fundamental as Yngve portrays it. The fact that linguists' theories so often nowadays seem to be "maintained by philosophical argument, polemics, or social pressure" more than by empirical evidence is a consequence of the historical accident that the discipline has allowed itself to be dominated by a few very powerful personalities who happen not to be strongly attached to the scientific ethos (cf. Pullum 1996). Linguistics could recover from this malady without needing to give up everything it knows about utterances, nouns, relative clauses, and so forth. I do not believe linguistics can ever be very similar to physics; but, since Yngve recognizes the importance of a long historical perspective in thinking about the problems of scientific method, I should like to commend one of the wisest observations I know on that topic:

> It is a mark of the trained mind never to expect more precision in the treatment of any subject than the nature of that subject permits. (Aristotle, *Nicomachean Ethics* 1094b 24–5)

References

Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett. 1974. *The Psychology of Language*. McGraw-Hill, New York.

Pullum, Geoffrey K. 1996. Nostalgic views from Building 20. *Journal of Linguistics*, 32:137–147.

Ruse, Michael. 1973. *The Philosophy of Biology*. Hutchinson, London.

Sampson, Geoffrey. 1997. Depth in English grammar. *Journal of Linguistics*, 33: 131–151.

Yngve, Victor H. 1960. A model and an hypothesis for language structure.

Proceedings of the American Philosophical Society, 104: 444–466.

Yngve, Victor H. 1961. The depth hypothesis. In Roman Jakobson, editor, *Structure of Language and its Mathematical Aspects*. American Mathematical Society (Providence, Rhode Island), pages 130–138. Reprinted in Fred W. Householder, editor, *Syntactic Theory I: Structuralist*. Penguin (Harmondsworth, Mddx.), 1972, pages 115–123.

Geoffrey Sampson is Reader in Computer Science and Artificial Intelligence at the University of Sussex; he was previously Founder Director of the Centre for Computer Analysis of Language and Speech at the University of Leeds. His recent books include *English for the Computer* (1995), *Evolutionary Language Understanding* (1996), and *Educating Eve* (1997). His web site is http://www.grs.u-net.com. Sampson's address is: School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton BN1 9QH, U.K.; e-mail: geoffs@cogs.susx.ac.uk