

## Clues from the Depth Hypothesis: A Reply to Geoffrey Sampson's Review

Victor H. Yngve\*

(The University of Chicago)

*In linguistics it has not been possible to use the standard criteria and assumptions of science because the ancients placed our discipline not in the physical domain but in the logical domain where concepts and theories do not represent parts of the natural world. Many of the problems facing linguistics follow inevitably, for example the difficulties that linguistics experiences in agreeing on grammatical theory. One symptom is the long-standing difficulty in testing the depth hypothesis, which came out of early MT research. Sampson (1997) attempted recently to test the depth hypothesis by a computer analysis of a grammatically annotated corpus of English. It is shown that this attempted test and his attempt at defending the testability of the depth hypothesis are invalid. But clues from the depth hypothesis have led to new foundations for general linguistics put forth in the book (Yngve 1996) that Sampson (1998) reviewed. This work reconstitutes linguistics in the physical domain where the criteria and assumptions of science can be applied. Sampson's review of this book contains a number of serious errors and inaccuracies.*

### 1. Introduction

It is touching that Geoffrey Sampson in his review (1998) of my recent book (Yngve 1996) should support the scientific integrity of the depth hypotheses in the face of my having shown that in its original formulation it cannot be tested. Thus I find myself in the curious position of having to argue against a review that strongly champions my own earlier work.

It is understandable that Sampson would be unhappy that I have walked away from the depth hypothesis in its original form since he has been building on it (Sampson 1997). His review thus largely ignores the deeper significance of the depth hypothesis also explored in the book. It has provided important clues leading to new foundations for general linguistics that open up exciting research opportunities, especially for computational linguists.

Sampson and I agree on the necessity of paying attention to empirical support for our statements and we agree that it is not scientifically legitimate to invent one's data. He characterizes corpus linguistics as a shelter beneath which scientific research can proceed undisturbed by work in areas "where people make their examples up out of their heads." There are problems with this cozy prospect, however. The shelter is subject to serious leaks. Perhaps the new foundations will provide a more secure shelter.

---

\* Preferred mailing address: 28 Crest Drive, Dune Acres, Chesterton, IN 46304. E-mail: v-yngve@uchicago.edu

## 2. Problems Testing the Depth Hypothesis

The depth hypothesis (Yngve 1960) came out of an early effort to test computer grammars by the random generation of sentences. The sentence-producing algorithm used phrase structure rules from a syntactic dictionary and expanded them from top to bottom and left to right so as to produce the output words sequentially in their normal order. It used a pushdown to keep track of the nodes in the developing syntactic tree structures. This raised the question of how much memory would have to be set aside for the pushdown so that it could handle the most complex structures it would ever encounter. An investigation showed that a pushdown holding only a few unexpanded nodes would be sufficient for producing even very long naturally occurring English sentences and even the most complex invented sentences would require no more than eight or nine unexpanded nodes. The structural correlate was an observed left-right asymmetry in occurring tree structures.

This surprising result led to an hypothesis that sentences of all languages are similarly constrained as a result of a limited human temporary memory, the famous  $7 \pm 2$  of George Miller. The hypothesis proposed to add to the known mechanisms of language change a mechanism involving occasional pushdown overload and consequent failure to remember what one was going to say. Speakers would then be led to tag certain syntactic structures involved directly or indirectly in pushdown overload as awkward. When awkward constructions were then avoided by speakers, children would not learn them and language change would occur in a direction favoring the observed tightly constrained structures. When alternative constructions involving fewer unexpanded nodes were substituted for the awkward ones, syntactic complexity in the grammar would be the expected result. In fact, it became possible to explain the utility of much of the observed syntactic complexity in English as an adaptation of grammar to the limited temporary memory available to speakers.

Some linguists tested the structural predictions of the hypothesis on various languages other than English and were led to support the hypothesis. But Chomsky opposed it not on the basis of any empirical tests but because he drew syntactic tree structures differently from the way I was drawing them. A rather confusing literature then grew up surrounding the depth hypothesis and efforts to test it.

As I tried to revise the sentence-producing algorithm (which had begun life as simply a module in a proposed machine translation scheme) into something that could be scientifically justified, it gradually became clear that lacking a proper model of how people speak and understand, all methods of drawing syntactic tree structures were equally arbitrary—mine, Chomsky's, and all others. There was simply no scientific way to choose among them. So I gradually came to realize that since the depth hypothesis was an hypothesis about syntactic tree structures and linguists could not agree on how to draw tree structures, the hypothesis as originally formulated would have to be rejected as untestable.

Curiously, even as I was gradually being driven to the realization that the hypothesis was untestable, some of its predictions were partially confirmed. The predicted forgetting phenomena were observed (Yngve 1973) and a review (Yngve 1975) of historical data from Charles C. Fries (1938, 1940) showed changes in word order in English from about 900 to 1300 A.D. that closely followed the predictions of the hypothesis. This only added to the puzzle. How could it be that a scientifically unacceptable hypothesis could appear to be supported by the observation of such predicted phenomena? However, these observations reinforced the idea that at least some aspects of the depth hypothesis may have a basis in fact. This kept alive the feeling that

the depth hypothesis could provide us with important clues as to how to proceed, a feeling that proved to be correct.

Recently, Sampson (1997) has tried to test the depth hypothesis by means of a careful computer investigation using the SUSANNE corpus, an impressive grammatically tagged corpus of English (Sampson 1995). This work is perhaps the most thorough attempt to test the depth hypothesis on naturally occurring text. The results did indeed show a left-right asymmetry in assumed tree structures. But gratifying as this result may be for champions of the depth hypothesis, it suffers from the same serious flaws as all other attempts to test it, including mine: disputable identification of tree structures that cannot be scientifically justified.

The SUSANNE corpus has two components: a transcription of naturally occurring English written texts and an annotation of the texts using what is described as "a comprehensive annotation scheme for representing the structure of the modern English language" (Sampson 1995, p. 1). The annotation scheme was set up arbitrarily by a group of researchers on the basis of rather vague criteria to provide what they felt to be an appropriate target for automatic parsing programs (Sampson 1995, p. 4). It is quite a stretch to see it as representing "the structure of the modern English language." A different group of researchers operating with different grammatical assumptions might well come up with a different annotation scheme incorporating a different way of drawing trees.

In fact, the way of drawing trees assumed in the SUSANNE scheme is different in important respects from that assumed in the original depth-hypothesis paper (Yngve 1960). Furthermore, in the thirty-five intervening years there have been many other grammatical schemes involving different ways of drawing trees and no end is in sight. The criteria are arbitrary inventions of the researchers or attempts to conform to intuition or to tradition or to some consensus. This has been the usual practice, but it is not science. There are no scientific criteria for deciding what grammatical scheme might be best or even scientifically acceptable.

In his attempt to test the depth hypothesis, Sampson made use of the parsings of the corpus done according to the SUSANNE scheme. The data thus reflect not only what the original writers wrote, but how the texts were analyzed on the basis of those debatable grammatical assumptions including how to draw trees. The arbitrary decisions in the annotation scheme, unsupported by any scientific evidence, have thus mixed with and contaminated the data. A computer investigation of the tree structures found in the annotated corpus therefore does not amount to a valid scientific test of the depth hypothesis. Nor does it measure "the precise constraint in English." In science, one cannot simply invent one's data, even in part. Everything must be justified on the basis of the evidence, and no scientific evidence is given to justify the details of the annotation scheme. The fact that he could not have anticipated the result he found does not validate the work as empirical scientific research; any nonsensical result would be equally unanticipated.

I am sorry that Sampson has stumbled into this pitfall that I fell into in 1960 and have only recently been able to climb out of. I hope he will be able to follow me out.

### 3. Two Conflicting Traditions

There were also many other problems with linguistic theory that were showing up in our attempts to find out how to translate languages. These problems even now give rise to nagging worries for all serious linguists, so much so that grammar is regularly reworked. And it is reworked again and again. The source of the difficulties lies in our acceptance of the semiotic-grammatical tradition from the ancient Greeks and

trying to adapt it to modern science. With its origins in ancient philosophy and logic, this tradition is simply incompatible with modern science. It gives us no choice but to invent or adopt arbitrary grammatical schemes through which to view the data. This is one of the compelling reasons why linguistics must now move to properly scientifically justified foundations.

Sampson states that "the fact that linguistic terminology originated in logical discourse does nothing in itself to establish the unscientific status of modern linguistics." This reflects two confusions. First, it is more than just terminology that is at issue: it is a question of whether to accept scientific or nonscientific foundations for linguistics. Second, the burden of proof in science is just the opposite. Science starts not with credulity or belief, but with doubt. The fact that linguistic theories descend from an ancient tradition that even today is widely followed and believed does nothing to establish their scientific integrity. The question really is, How do we decide what to believe in linguistics? I think Sampson would agree that we must pay attention to empirical support for our theories. Conformity to a tradition, no matter how hallowed, does not automatically imply adequate empirical support.

Sampson's analogy with astronomy is appropriate if understood in the context of moving from an ancient tradition into modern science. Four centuries ago, people believed that the orbits of the planets must be perfect circles because this was a perfect geometrical figure. The way in which this erroneous theory was replaced by more scientifically acceptable theories is instructive. Tycho Brahe had made the most precise observations of the positions and motions of the stars and planets possible before the use of the telescope. Kepler undertook to calculate from these observations what the orbit of Mars actually was. He found that it was an ellipse. He deserves great credit for going with the data and against the ancient tradition. Another good example is Galileo, who argued that the moon was not a perfectly smooth sphere, as the ancient tradition would have it. He went with the data he obtained with his telescope where he actually calculated the heights of the mountains on the moon from the changes in their shadows as the angle of the sunlight illuminating them changed.

So we agree that linguistics should be more empirical. I would rather say that it should be properly scientific, for science involves more than empiricism. One cannot study objects scientifically that have no observable reality. Yet Sampson appears to be under the impression that a grammatical linguistics focused on language could be scientific if only it were to pay more attention to empirical studies of grammatically annotated corpora. The burden of proof for this belief is on those who wish to continue to hold it. His attempted empirical test of the depth hypothesis does not contribute to such a proof.

I can assure Sampson that I am as disappointed as he must be that the depth hypothesis as originally formulated is untestable, that it cannot stand as a part of standard science, but only as pointing to interesting phenomena for which we have as yet no proper scientific explanation. This realization did not come easily to me because of my personal bias and the strength of the semiotic-grammatical tradition. I would not expect it to come easily to Sampson or others who have also invested time and effort in trying to test it. But it is something that we must face up to and deal with honestly. So where should we turn? The clues as to how to proceed proved to be embedded in the depth hypothesis.

The depth hypothesis was much more than simply an hypothesis about a past-future asymmetry in tree structures. It was about a model that proposed the underlying physical reasons for such an asymmetry in the data in terms of the operation of a sentence-producing mechanism with a pushdown using a limited temporary memory. It had components related to forgetting what one was about to say, points about

feelings of awkwardness, points about learning, and points about linguistic variation and historical change.

I do still believe that there is a grain of truth in the depth hypothesis. This grain of truth is related in the degree to which the sentence-producing model was a model, however imperfect, of how people speak. The problems with the hypothesis are related to the degree to which it accepted a number of scientifically unsupported assumptions from the semiotic-grammatical tradition. The depth hypothesis is providing us here with two clues. It is telling us that linguistics would be better off investigating people, a physical reality, rather than to continue trying to investigate language, which, after all, has no physical reality and is simply introduced by assumption. It is also telling us that we should follow science rather than the semiotic-grammatical tradition, which began in philosophy and has never been scientific.

#### 4. Standard Science and the New Foundations

Although Galileo, Kepler, and their contemporaries faced great hurdles in moving astronomy into science, they had one advantage that linguistics does not share. The ancients had already conceived of astronomical objects as in the physical domain where they were available to observation by the senses and thus potentially amenable to scientific study. But the ancients conceived of linguistic phenomena as in the logical domain where objects are merely assumed and thus unobservable and not amenable to scientific study. Although it is real observable people in the physical domain who talk and understand, the semiotic-grammatical tradition they developed has not provided us with even initial linguistic concepts and theories in the physical domain where scientific investigations could potentially be carried out. Before we can do standard science in linguistics we must reconstitute our discipline in the physical domain. My book does exactly this: it provides new, scientifically acceptable foundations for general linguistics focused on people in the physical domain.

The most far-reaching change we face in moving fully into modern science is in the standard criteria and assumptions developed over four centuries and universally accepted in science. It is worth going over them briefly here; they are treated in detail in the book.

Science offers standard criteria for what to believe about the natural world:

- The standard criterion of acceptance of hypotheses in science when doubts arise is the ability of their predictions to pass tests against the real world by means of careful observations and experiments.
- The standard criterion of acceptance of observational and experimental results in science is their reproducibility when questioned.

This means that we turn our back on all other criteria such as conformity to tradition, conventional wisdom, fad, personal whim or intuition, or allegiance to a teacher, a charismatic individual, or a school of thought. A colleague recently remarked in the question period following a talk on the criteria of science and the new foundations, "I agree with everything you say, but I like what I'm doing." Such a criterion (personal pleasure) does not belong in science.

Modern science has advanced by starting with doubt. It has questioned all its assumptions, traditional or modern, and has rejected all that do not stand up to the above

criteria except the following four, which have become the four standard assumptions of all science:

1. An ontological assumption: that there actually is a real world out there to be studied.
2. A regularity assumption: that the real world is coherent, so we have a chance of finding out something about it.
3. A rationality assumption: that we can reach valid conclusions by reasoning from valid premises, that we can trust our ability to calculate predictions from our theories for comparison with the real world.
4. A causality assumption: that observed effects flow from immediate real-world causes.

Note particularly that according to the standard assumptions, science studies the real world, that it proposes causal theories that model the real world, that it then tests these theories by comparing their predictions with observations and experiments carried out on the real world.

In the process of rejecting all assumptions but these four, science has rejected much ancient myth, folk theory, and superstition. In the case of linguistics, this means we must reject all the assumptions of the semiotic-grammatical tradition including the assumptions that introduce language and the objects of language. We must also reject certain assumptions inherited or acquired from philosophy. Accepting any of them will jeopardize the scientific integrity of the discipline.

The new foundations focus on people, sound waves, and other real objects that may be involved in communicative behavior. Theories on the new foundations include dynamic causal models of people and how they talk and understand and communicate in other ways. These are not models of how people “use language” because that would involve unsupported assumptions from the semiotic-grammatical tradition. There are also dynamic causal models of groups of people and the other physical concomitants of communicative behavior. These foundations are broad enough and strong enough to accommodate all linguistic phenomena in all branches of general linguistics including also developmental, variational, and historical phenomena and nonverbal communicative behavior and writing. The new foundations include a scientifically justified computable notation for representing our theories and calculating their predictions for comparison with observation.

This does not mean that we must discard all we have ever found out in linguistics. It does mean, however, that we can take none of it for granted. Everything must be carefully reexamined and any grains of truth sifted out.

The dictates of the scientific tradition are severe and demanding if one wishes not to publish nonsense. The new foundations for general linguistics laid out in the book are the end results of a decades-long effort. Earlier discussion is to be found in my 1986 book and a series of papers in the *LACUS Forum* volumes. I believe that the new foundations are scientifically sound all the way down to bedrock.

## 5. Implications for Computational Linguistics

My 1996 book that Sampson reviewed restricts itself to laying new scientifically justified foundations for linguistics. It does not attempt to build a new linguistics on these foundations. That would require a sustained program of research. Sampson does not

seem to understand this in spite of discussion throughout the book, especially in the introduction to Chapter 21 and even the book's subtitle, *New Foundations for General Linguistics*.

The book does, however, survey a number of areas of linguistics and discusses the directions that new theory-building might take and how we can obtain scientifically acceptable evidence on a number of fronts. We need to build on the new foundations testable models of people and groups and how they talk and understand and communicate in other ways, and then we need to test the models.

Here is where computational linguistics can make its particular contribution. The new foundations provide a computable notation for representing linguistic theories of people and groups that is scientifically justified in its foundations rather than being introduced a priori or on the basis of scientifically unsupported assumptions. This means that unlike the depth hypothesis, our models have a chance of being scientifically testable, that they have a potential for leading us closer to the truth about what is perhaps most characteristically human about humans, how they communicate.

Understanding in context is a dynamic feature at the most basic level of models built on the new foundations. Our models will understand what they are doing in the most fundamental sense, not in any grammatical, semiotic, or logical sense, which would not accurately reflect what is going on physically.

From the tenor of Sampson's review, I don't think he has yet understood that linguistics actually can develop as a science like the other sciences, that the problem is not just a lack of scientific ethos among certain leading linguists. If the reader is to learn anything from the review it is that the book needs to be read carefully and studied chapter by chapter as a textbook rather than being skimmed like a novel or even the description of yet another form of grammar; skimming inevitably leads one to read into a work what one expects rather than what is actually there.

There are a number of additional errors and misconceptions in Sampson's review. I will mention some of them here.

The review misunderstands the simplified illustrative examples devised to explain the new conceptual structure for attempts at building on the new foundations. They are illustrative of the theory and the notation, not finished results. Thus the comments on this basis are ill-conceived and erroneous.

I don't believe that Sampson understands the breadth of data available to linguistics on the new foundation. This is treated throughout and is the special topic of whole chapters. He seems not to have understood the distinction I make everywhere between observation and theory and between theory and the real world. He seems to think that I admit only the data of sound waves. I do not imply anywhere that the relation between speech and orthography is not complex nor do I imply that one can relate airwaves directly to observable behavior. Sampson does not seem to have grasped the hierarchical structure of theory on the new foundations.

Sampson complains that I sometimes represent what people are saying in my illustrative examples in terms of ordinary orthography. He is mistaken if he thinks this means I condone any grammatical scheme for analyzing a corpus. He has apparently completely missed the treatment of writing and written text in Sections 17.8 and 22.6 and elsewhere. See also the treatment of writing and written text in Yngve (1991).

Contrary to the misrepresentation in the review, models on the new foundations are not confined to binary variables if the nature of the evidence does not warrant them. I do not imply that "it" in tag and "on the spot" to answer a question are "theory-free predicates." I place them squarely in the realm of theory and discuss at some length the evidence that one could obtain concerning them. And there is nothing in the book that implies that the models cannot permit more than a finite number of

combinations. Sampson's discussion of the problems in evolutionary biology and his quotation from Aristotle are beside the point.

Sampson apparently hoped to find finished analyses in the book that he could then compare with analyses in the grammatical tradition. Interested readers will find a seventeen-point comparison with other approaches in Yngve (in press).

### References

- Fries, Charles C. 1938. Some notes on the inflected genitive in present-day English. *Language*, 14:121–133.
- Fries, Charles C. 1940. On the development of the structured use of word-order in modern English. *Language*, 16:199–208.
- Sampson, Geoffrey. 1995. *English for the Computer: The SUSANNE Corpus and Analytic Scheme*. Clarendon Press, Oxford, UK.
- Sampson, Geoffrey. 1997. Depth in English grammar. *Journal of Linguistics*, 33:131–151.
- Sampson, Geoffrey. 1998. Review of *From Grammar to Science: New Foundations for General Linguistics* by Victor H. Yngve. *Computational Linguistics*, 24(1):173–176.
- Yngve, Victor H. 1960. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104:444–466.
- Yngve, Victor H. 1973. I forget what I was going to say. In Claudia Corum et al., editors, *Papers from the Ninth Regional Meeting, Chicago Linguistic Society*, pages 688–699, Chicago Linguistic Society, Chicago.
- Yngve, Victor H. 1975. Depth and the historical change of the English genitive. *Journal of English Linguistics*, 9:47–57.
- Yngve, Victor H. 1986. *Linguistics as a Science*. Indiana University Press, Bloomington and Indianapolis.
- Yngve, Victor H. 1991. Concepts of text and knowledge. In Angela Della Volpe, editor, *The Seventeenth LACUS Forum, 1990*, pages 539–550, Linguistic Association of Canada and the United States, Lake Bluff, IL.
- Yngve, Victor H. 1996. *From Grammar to Science: New Foundations for General Linguistics*. John Benjamins, Amsterdam and Philadelphia.
- Yngve, Victor H. In press. Two Foundations for Linguistics Briefly Compared. To appear in *The 24th LACUS Forum, 1997*, Linguistic Association of Canada and the United States, Lake Bluff, IL. Preprint available at <http://humanities.uchicago.edu/humanities/linguistics/faculty/ynovelacus.html>