

# Letter to the Editor

## Grammatical Depth: A Rejoinder

Geoffrey Sampson<sup>1</sup>  
(University of Sussex)

I was a little startled to find Victor Yngve (1998) using my review of his book *From Grammar to Science* as a peg on which to hang a sustained critique of my article “Depth in English grammar” (Sampson 1997), an item that was mentioned only tangentially in the book review.

My 1997 article tested Yngve’s famous depth hypothesis of forty years ago against data of a kind richer than was available then. It showed that the hypothesis as formulated by Yngve was not precisely the correct generalization about English, but it also showed that there does exist an invariant quantitative property of English syntactic structures that is closely related to Yngve’s hypothesis.

Yngve regards this demonstration as empty, on the grounds that there are different schemes for representing English grammatical structure, and hence my finding could have been merely the result of an arbitrary choice of analyses. “The fact that [Sampson] could not have anticipated the result he found does not validate the work as empirical scientific research; any nonsensical result would be equally unanticipated” (Yngve 1998, p. 635).

Of course there are alternative schemes of grammatical analysis. My article drew attention to this, and surmised that the finding might be robust with respect to choice of analytic scheme. (This surmise could prove mistaken, but Yngve does nothing to suggest that it was in fact mistaken.) The point that seems to be lost on Yngve, though, is that there was no a priori reason to expect the data to yield *any* result as specific and precise as the result that emerged. It was a surprise to me to find, long after publication of the SUSANNE Corpus, that its structures displayed a quantitative property whose distribution possesses so low a standard deviation round its mean as does the property of “raw production-based sentence depth” defined in Sampson (1997).

Yngve seems to imagine that the team responsible for the SUSANNE Corpus brought about this outcome by making arbitrary analytic decisions. Protestations about our research ethics would be redundant here; we would not have known how to cook the books that way even if we wanted to. No member of the team that produced the SUSANNE Corpus was aware of the measure that later turned out to be highly invariant. (The team had scattered by the time I engaged in the research reported in Sampson [1997], so most members are very likely unaware of it even now.) If we had been conscious of it, I cannot imagine how we could have gone about forcing our analyses to conform to the quantitative invariant, while achieving consistency with published analytic guidelines defined in great detail and in entirely nonquantitative terms. All that effort, just to manufacture a basis for one 20-page journal article? I don’t think so. The invariant was there in the language samples; we didn’t put it there.

On the wider issue, whether empirical science can—never mind should—be

---

<sup>1</sup> Cognitive and Computing Sciences, University of Sussex, Brighton BN1 9QH, U.K. E-mail: geoffs@cogs.sussex.ac.uk

founded on absolutely reliable observation statements, I have already expressed my differences with Yngve in my review of his book. Yngve makes his view very explicit when he claims (1998, p. 638) that his "new foundations" for linguistics "are scientifically sound all the way down to bedrock." I believe, with Sir Karl Popper, that "science does not rest upon solid bedrock" (Popper 1959, p. 111). Rather, in Popper's figure, scientific theories are structures erected on piles driven into a swamp. If any particular support seems unsatisfactory, it can be driven deeper; but there is no ultimate bottom to be reached anywhere.

#### References

- Popper, Karl R. 1959. *The Logic of Scientific Discovery*. Hutchinson, London.
- Sampson, Geoffrey R. 1997. "Depth in English grammar." *Journal of Linguistics* 33:131–151.
- Yngve, Victor H. 1998. "Clues from the depth hypothesis: A reply to Geoffrey Sampson's review." *Computational Linguistics* 24(4):633–640.