ON SOME AXIOMS ABOUT SENTENCE LENGTH

D. L. Olmsted

University of California, Davis

Some postulates concerning the length of sentences, as advanced by Noam Chomsky, are considered in the light not only of grammatical possibilities, but also of behavioral probabilities; and certain implications for linguistic theory are suggested.

In his *Syntactic structures*, Chomsky makes the following statement (23): 'We might arbitrarily decree that such processes of sentence formation in English as those we are discussing cannot be carried out more than n times, for some fixed n. This would of course make English a finite state language, as, for example, would a limitation of English sentences to length of less than a million words.'

I propose to discuss this passage, not with the idea of knocking out one of the essential props of Chomsky's system (since in the remarks that immediately follow, Chomsky makes it clear that he has serious objections to finite state grammars even without these assumptions), but in order to highlight some differences in approach that significantly affect the formation of axioms.

These differences result from the fact that some linguists approach the study of language in fieldwork situations, where they must analyze languages previously unknown to them, whereas others make their approach from a background in symbolic logic and in contexts which include programming computers to aid in analysis or translation.

That the aims of the two groups of investigators are not identical is suggested by the following passage from *Syntactic structures* (56):

In short, we shall never consider the question of how one might have arrived at the grammar whose simplicity is being determined; e.g. how one might have discovered the analysis of the verb phrase presented in 5.3. Questions of this sort are not relevant to the program of research that we have outlined above. One may arrive at a grammar by intuition, guesswork, all sorts of partial methodological hints, reliance on past experience, etc. It is no doubt possible to give an organized account of many useful procedures of analysis, but it is questionable whether these can be formulated rigorously, exhaustively, and simply enough to qualify as a practical and mechanical discovery procedure.

But of course the field worker, approaching a language unknown to him, finds the question of how to arrive at the grammar not only relevant, but all-important. He wants explicit instructions on the construction of the grammar, and has traditionally been willing to leave the comparison of grammars to be decided by intuition.[1] Chomsky's contribution has been to show that it is possible to devise rigorous procedures for the evaluation of grammars, but he has been willing to leave procedures for arriving at the grammar to intuition. A linguistics that is both scientific and complete will have to eliminate, at least in principle, intuition as a decision procedure in both these areas of investigation.

Both approaches have contributed to linguistics, and ultimately our science

---

[1] Notable exceptions in the forties and early fifties were Z. S. Harris, C. F. Hockett, R. S. Wells, and C. E. Bazell.

will have to rest on a set of postulates useful in both contexts. At present, however, each approach makes use of postulates not employed—and frequently rejected—by practitioners in the other context. I have been using 'approach' and 'context' rather than 'theory' to emphasize the belief that the two groups of investigators are dealing with crucially different (even though apparently overlapping) sets of data, rather than presenting competing theories for the explanation of the same set of data. The different sets of data are selections viewed by the investigator as relevant to his interests. These choices of data, in their turn, influence the assumptions of the linguist as he begins his study.

A case in point is the relatively casual dismissal of the 'arbitrary limitation' of English sentences to length of less than a million words. The assumption that English sentences are not always less than a million words in length smells of the computer center rather than of the field situation. The field worker is quite ready to assume the contrary. Indeed, it may be stated as an empirical generalization that, using any definitions of 'sentence' and 'word' so far found useful in describing actual utterances, every sentence in every language so far studied is less than one million words long. That generalization is, I believe, uncontradicted in the empirical literature of linguistics.

Why, then, should Chomsky make such an assumption? It is apparent from what follows the quoted passage that at least one of his reasons is the desire to retain recursive devices in his grammar and to avoid fixing n at any particular value—because 'if these processes have no finite limit, we can prove the literal inapplicability of this elementary theory' (23). Possibly the finite state theory can be proved inapplicable; but if it is desirable to do so, the proof ought to be made to rest on stronger foundations than those contained in the axiom about sentence length.

Though we may be willing to concede that the apparently unlimited recursiveness of English sentences gives the infinite length axiom a certain GRAMMATICAL utility, there are a number of reasons to view it with suspicion. One such reason is the fact that sentences consist of more than words (however defined): there are also intonation patterns. These latter can be ignored by anyone using a machine to analyze sequences of words, but not by the field worker. For him they are important clues, as they presumably are to the ordinary hearer. The use of a sentence-final intonation pattern is probably physiologically independent of the change points between expiration and inspiration, but it is a good bet that there is a cultural tendency for there to be a coincidence between the two. That is, it may be stated as a theorem that a sentence-terminal pattern will coincide with the end of a breath group at least once in every n breaths, where n is a digit greater than zero.

There is a more fundamental behavioral issue at stake: the function of language as a means of communication. To set out to generate a string a million words long, without any sentence-final intonation pattern, implies no chance for any possible hearer to reply, and hence implies disinterest in (or rejection of) any possible reply. This is not only not communication in any usual human sense, but such a speaker would probably be committed as a raving madman, (no matter how 'grammatical' his utterance) after, say, a few days of talking

without ending a sentence. Indeed, performances along this line that fall far short of a million words constitute well-known diagnostic marks of hysteria, the manic phase of manic-depressive psychosis, and other maladies. The cross-cultural literature on shamanistic practice and mental illness, though not as complete as might be wished, makes it clear that this sort of dysfunction of linguistic communication can be recognized quite without reference to the content of the speech. Of course, the examples all fall far short of a million words.

How far short? The answer will be of considerable importance in the eventual determination of Chomsky's n. Determining n will be messy, and I realize that Chomsky probably doesn't want to determine it anyway. It seems probable that a speaker has to keep n quite low in order to be understood, even when he makes full use of the intonation patterns that aid in the perception of recursive sequences.

Thus the assumption of infinite sentence length is not only not factual but, what may seem more important to some, it is counter-intuitive. Probably the assumption can be abandoned without any great damage to Chomsky's system. His ideas, which have stimulated the production of so much interesting work, will be even more valuable if they can be fitted to assumptions which will be as useful in the sweathouse as in the study.