

The Vastness of Natural Languages

D. TERENCE LANGENDOEN
and
PAUL M. POSTAL

Contents

<i>Preface</i>	vi
1 Set – Theoretical Background	1
2 The Received Position about NLs and their Grammars	14
3 Sentence Size Bounds	27
4 The Analogy with Cantor's Results	49
5 Implications	71
6 Ontological Escape Hatches	104
7 The Characterization of Transfinite Sentences	157
<i>References</i>	177
<i>Index</i>	185



Basil Blackwell

Preface

This monograph argues that the collection of sentences comprising each individual natural language (NL) is so vast that its magnitude is not given by any number, finite or transfinite. This means that NLs cannot, as is currently almost universally assumed, be considered recursively enumerable, hence countable (denumerable) collections of sentences. For if they were such, the magnitude of each would be no greater than the smallest transfinite cardinal number \aleph_0 . It then follows that there can be no procedure, algorithm, Turing machine or grammar that constructs or generates all the members of an NL, since, by definition, such a procedure, algorithm, Turing machine or grammar can construct or generate only recursively enumerable, hence countable, collections. A system which constructs some NL sentences must inevitably leave most NL sentences unconstructed.

Linguistic research over the past quarter century has been largely guided by two major assumptions introduced by N. Chomsky: (i) that the best theory of NLs is a theory of grammars that generates NLs and (ii) that human beings know an NL in virtue of knowing a grammar that generates it. These assumptions cannot be maintained. The only adequate theories of NLs are, we contend, those that posit non-constructive or non-generative grammars. The theorems of such theories are truths about NLs in general or about particular NLs; one of these theorems states that no procedure, algorithm, Turing machine or grammar can construct or generate all the sentences of any NL. Accordingly, if human knowledge of NLs were limited to what internalized grammars can generate, then that knowledge would be extraordinarily deficient. Put differently, if psychologically real (internalized)

grammars are generative, then those grammars describe not NLs, but at best only proper subcollections of NLs.

If NLs are merely recursively enumerable collections of sentences, then Chomsky's two major assumptions are compelling. A grammar that generates the sentences of a recursively enumerable collection is, first, a complete theory of that collection in the sense that no sentence of the collection is left out. Second, to account for the remarkable ability of human beings to decide the grammatical properties of enormously many NL sentences, they may reasonably be supposed to have the ability to generate representations of those sentences; under ideal conditions, the procedures that underlie that ability could then be used to determine the grammatical properties of any NL sentence whatever. However, the appeal of Chomsky's assumptions should not obscure their dependence on the even more basic assumption that NLs are recursively enumerable (hence denumerable) collections. This assumption is, we maintain, false, and any implications drawn from it, no matter how appealing or widespread, are inevitably also at least partly in error.

This study had its origins in several strands of work: (i) efforts to make precise the notion generative grammar and the collections that such grammars generate (see Langendoen (1976, 1982)), (ii) the discovery that there are essential truths about NLs that are not expressible within any theory of generative grammar; (iii) efforts to develop a non-constructive theory of grammar (see Johnson and Postal (1980); Postal (1982)), and (iv) realization that the truths about NLs that are not expressible constructively are expressible non-constructively. Finally, this work has been influenced and inspired by Katz's work on the abstract character of NLs; see Katz (1977, 1978, 1980, and especially 1981, 1983). Although one does not have to accept platonic realism to accept the conclusions drawn here about the vastness of NLs, the two go very naturally together. Moreover, the only conceivable rational basis for rejecting our conclusions is the adoption of an ontological commitment (nominalism or some version of conceptualism) so narrow that it either does not recognize the existence of certain objects claimed here to be NL sentences, or else, without justification, excludes these objects from the proper realm of linguistics.

Our conclusion concerning the vastness of NLs is based on a demonstration of a strict parallelism between the collection of all

sentences of an NL and the collection of all sets. The discovery around the turn of this century that the latter collection is not itself a set led to fundamental reforms in logic and the foundations of mathematics. The same reasoning that establishes that the collection of all sets cannot itself be a set (a collection with fixed magnitude, finite or transfinite) also establishes that the collection of all NL sentences cannot be a set. Consequently, fundamental revisions of currently standard views of NLs and grammars are required for reasons similar to those that operated in the foundations of logic and mathematics. For readers who are unfamiliar with set theory or who wish to refresh their memories, chapter 1 provides a brief outline of the relevant aspects of set theory and a discussion of their relevance to linguistic issues.

Chapter 2 points out that the standard linguistic assumption that NLs are infinite recursively enumerable collections of sentences each of finite length (size) has been defended against the position that NLs are finite collections of sentences each shorter than some fixed finite length. However, practically no arguments have been given against the position that NLs are non-denumerable collections of sentences, some (in fact, most) of which are of transfinite length. Chapter 3 examines the arguments in the literature that purport to show that there can be no *finite* length limitation on NL sentences. Most of these arguments are unsound; however, one argument is not only sound but generalizes to show that there is no *transfinite* length limitation on NL sentences either.

Chapter 4 demonstrates that the classical proof that the collection of all sets itself is not a set also shows that the collection of all sentences of any NL is also not a set. This demonstration reveals a property of NLs that is inexpressible by any constructive or generative theory of NLs. Indeed, each generative theory of NLs entails that NLs do not have the property in question.

Chapter 5 draws out certain consequences for linguistic theory of the argument in chapter 4, taking up such matters as the incorrectness of almost all contemporary theories of NL grammars, the character of non-constructive grammars of NLs, and the effability and learnability of NLs. Chapter 6 considers whether the implications of the main result of this study can be avoided by adopting one or another non-platonist view about the nature of NLs. We contend that none of these ontologies offers any insight into the

nature of NLs nor any insight into the relation of NLs to human beings or to human knowledge of NLs that is not available within the framework proposed here. Since each of these alternatives requires arbitrary imposition of a restriction on the size of NL sentences otherwise not required, the conclusions drawn about the vastness of NLs are shown to be unavoidable. Finally, chapter 7 considers the issue of transfinite sentences from a different point of view, arguing that any approach which attempts to place them beyond the bounds of linguistic study yields absurdities. It also takes up the question of the essential unity of the realms of finite and transfinite sentences for each NL.

We would like to offer our special thanks to Jerrold Katz and Arnold Koslow for helpful discussions of the issues involved in this study. We are also grateful to David Johnson, Edward Keenan and Warren Plath for many useful comments on earlier versions of this work. Any remaining errors are ours, not theirs.