

Preprint No. 12

Classification: CA 1.2

Quantitative Approximation to the Word

- E. Gammon -

INTERNATIONAL CONFERENCE

ON

C**OMPUTATIONAL L***INGUISTICS



C O L I N G

1969



RESEARCH GROUP FOR QUANTITATIVE LINGUISTICS

Address: Fack Stockholm 40, SWEDEN

QUANTITATIVE APPROXIMATIONS TO THE WORD

Edward Gammon

California State College at Fresno, USA

INTRODUCTION

In this paper I want to distinguish between the behavior of language and the behavior of linguists in analysing linguistic data. The paper concerns distributional approximations to the word as a sequence of sub-word units. I will argue on the basis of what a hypothetical linguist might do with hypothetical linguistic data-- in this case the frequencies of sequences of morphemes. The behavior of the hypothetical linguist is formalized into a quantitative model, which in turn is applied to real linguistic data: a primer, a lower school reader, and a well known tract on the philosophy of language.

The model that we arrive at is a particularly simple formulation of a discovery procedure for words. This formulation is part of my longer concern with the aetiology of ideas about linguistic units: why does a linguist feel a phrase is a phrase and a word is a word?

The Word as a Distributional Unit

The status of the word as a linguistic unit is best summarized by Greenberg:(1957)

The word as a unit occupies a paradoxical position in contemporary linguistic science. Such a unit, roughly coinciding in usage with its employment in everyday language and in the discourses of sciences other than linguistics, occurs almost universally in practice of descriptive linguists as the dividing line between the two levels of morphological (intra-word) and syntactic (supra-word) constructions. Yet no generally accepted