

prise about 355,000 entries selected from an unabridged dictionary and four technical dictionaries. The lists have been prepared by alphabetizing- and merging-routines on a UNIVAC I computer and reproduced on paper in 8-volume sets of limited supply. Tape version duplications are being provided on UNIVAC I and IBM tapes. More extensive listings are being prepared by the inclusion of entries from several more specialized dictionaries and glossaries.

Phonemicized versions of syllabic subsets are being developed on the basis of the present orthographic lists, and will be used in studies of the voids and predictabilities within the phonological rules for syllabic subsets, and in studies of mechanical recognition of morpheme and word boundaries. Research is also being carried out on word-derivational processes and the transformational equivalents of compound structures.

Support is provided by the U. S. Air Force Office of Scientific Research.

5.3.37 UNIVERSITY OF PENNSYLVANIA

Department of Linguistics, Philadelphia, Pa.

ZELLIG S. HARRIS, *Professor of Linguistics*

The computable analysis of language structure and the detailed investigation of linguistic transformations continue. It is expected that, in addition to its theoretical interest, this work may have a new kind of application to information retrieval. The project has resulted so far in a large body of syntactic studies on English, and in a working UNIVAC program which analyzes a syntactic (constituent) structure of English sentences (without the assistance of any human editing). At present, a transformation program capable of reducing any English sentence to component sentences is being put on a computer. In addition, work is being done on the detailed English transformations, on a general theory of transformations, and on the transformations of other languages.

The project is supported by the National Science Foundation.

References:

- (1) The main results of the project are given in Papers on Formal Linguistics, a series of monographs published by Mouton and Co., The Hague, Netherlands. Those published or in press:
PoFL 1. *String Analysis of Sentence Structure*, by Zellig S. Harris, 1962. (A revised version of TDAP 15, below)
PoFL 2. *Discourse Analysis Reprints*, by Zellig S. Harris. (In press)

(2) More detailed or tentative presentations of results are given in the series, Transformations and Discourse Analysis Papers (TDAP). The major numbered reports to the National Science Foundation to date are:

2. *Introduction to Transformations* (Reprint, 1956), Zellig S. Harris.

3, 4. *Discourse Analysis Reprints*, Zellig S. Harris.

5. *Generation of the Nominal Compound in English*, by R. B. Lees. This paper was later expanded into a thesis which was published as "The Grammar of English Nominalizations," *International Journal of American Linguistics*, Part II, vol. 26, no. 3, July 1960. Available from Director of Publications of the Research Center in Anthropology, Folklore, and Linguistics, Indiana University, Bloomington, Ind. Price: \$4.00.

11. *Preparing a Text for Retrieval* (Tentative), Zellig S. Harris.

12. *An Effective Procedure for Mechanical Preparation of a Scientific Text for Information Retrieval*, Ruth Zeitlin Koenigsberg.

14. *A Computer Program for Syntactic Analysis*, Zellig S. Harris.

15. *Computable Syntactic Analysis*, Zellig S. Harris. (Revised version published as PoFL 1, above)

16. *Word and Word-Complex Dictionaries*, Lila Gleitman.

17. *Elimination of Alternative Classifications*, Naomi Sager.

18. *Recognition of Local Substrings*, Aravind K. Joshi.

19. *Higher-order Substrings and Well-formedness*, Bruria Kaufman.

20. *Iterative Computation of String Nesting (Fortran Code)*, Bruria Kaufman.

21. *Syntactic Completion Analysis and Theories of Grammatical Categories*, Henry Hiz.

22. *Pre-adjectivals in the English Nominal Phrase*, Seymour Chatman.

27. *Procedure for Left-to-Right Recognition of Sentence Structure*, Naomi Sager.

28. *Problems in the Computation of English Sentence Structure (Substring analysis)*, Naomi Sager.

29. *The Intuitions of Grammatical Categories*, Henry Hiz.

30. *English Transformation List*, Zellig S. Harris.

31. *Order of Adjectives*, Zeno Vendler.

32. *A Procedure for Decomposing a Complex Sentence Into Its Kernels*, Aravind K. Joshi.
33. *A Decision Method for a Class of Formal Systems, and its Application to Transformational Analysis*, Hans Herzberger.
34. *Questions and Answers*, H. Hiz.
35. *The Classification of English Verbs by (Non-Sentential) Object Types*, Seymour Chatman.
38. *The Transformational Status of the Definite Articles in English*, Beverly Robbins.
39. *An Analysis of Container Sentences*, Ruth Z. Koenigsberg.
40. *Conjunction with and*, Lila R. Gleitman.
41. *Transformational Decomposition Systems*, Hans Herzberger.
42. *A Procedure for a Transformational Decomposition of a Complex Sentence*, Aravind K. Joshi.
44. *On Some Connections Between Transformations*, Zbigniew Lis.

Paper 15 gives an information presentation of a general theory and method for syntactic recognition. Papers 16-19 give the actual flow charts of each section of the syntactic analysis program.

- (3) Harris, Zellig S. "Linguistic Transformations for Information Retrieval," in *Proceedings of the International Conference on Scientific Information*, pp. 937-950. Washington, D. C.: National Academy of Sciences-National Research Council, 1959.
- (4) Hiz, Henry. "Steps Toward Grammatical Recognition," in *Advances in Documentation and Library Science*, vol. 3, part 2, pp. 811-822. New York-London: Interscience Publishers, Inc., 1961.
- (5) Joshi, Aravind K. "Computation of Syntactic Structure" (TDAP No. 25), in *Advances in Documentation and Library Science*, vol. 3, part 2, pp. 831-840. New York-London: Interscience Publishers, Inc., 1961.
- (6) Gleitman, Lila. "The Isolation of Elements for Grammatical Analysis," in *Advances in Documentation and Library Science*, vol. 3, part 2, pp. 823-830. New York-London: Interscience Publishers, Inc., 1961.
- (7) Chomsky, Noam. "On Certain Formal Properties of Grammars," *Information and Control*, vol. 2, 1959, pp. 137-167.
- (8) Hiz, Henry. "Congrammaticality, Batteries of Transformations and Grammatical Categories," in *Proceedings of the Symposium in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects*, ed. by Roman Jakobson, pp.

- 43-50. Providence, R. I.: American Mathematical Society, 1961.
- (9) Herzberger, Hans G. "The Joints of English," in *Proceedings of the Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects*, ed. by Roman Jakobson, pp. 99-103. Providence, R. I.: American Mathematical Society, 1961.
- (10) Goodman, Nelson. "Graphs for Linguistics," in *Proceedings of the Symposia in Applied Mathematics, Volume XII: Structure of Language and its Mathematical Aspects*, ed. by Roman Jakobson, pp. 51-55. Providence, R. I.: American Mathematical Society, 1961.
- (11) Hiz, Henry. "Grammatical Category as a Relation," in *Abstracts of Contributed Papers at the International Congress for Logic, Methodology and Philosophy of Science* (Stanford University, Stanford, Calif.), 1960, pp. 147-148.
- (12) Hiz, Henry. "The Intuitions of Grammatical Categories," *Methodos*, vol. 12, no. 48, 1960, pp. 1-9.
- (13) Hiz, Henry. "Modalities and Extended Systems," *The Journal of Philosophy*, vol. 58 (1961), pp. 723-731.
- (14) Hiz, Henry. "Questions and Answers," *The Journal of Philosophy*, vol. 58 (1962), pp. 253-265.

UNIVERSITY OF PITTSBURGH

5.3.38

*Computation and Data Processing Center,
817 Cathedral of Learning, Pittsburgh 13, Pa.*

DALE W. ISNER, *Principal Investigator*

Investigation underway concerns the feasibility of developing a programming language compatible with the needs of psychologists, educators, etc., in retrieval and analysis of natural text material. Existing languages in this field are being compared, and the requirements which such a language must have to be beneficial and also usable are being determined. The languages worked with are COMIT and Wegstein's String Manipulations in ALGOL. The IBM 7070/1401 system is used. Possible implementation on several computers is planned.

Thus far, samples of existing languages have been studied to determine their deficiencies, and various people using textual material as input to computers have been interviewed.

It is planned to evaluate existing languages and to innovate new ideas to determine specifications.

Partial support is provided by the International Business Machines Corporation.