

From the aspect of divisions, it is shown that Division I is quite unlike Divisions II and III. In Division I were included most of J and P, and in Division III most of E. E appears for the first time in Division II, and from there it has approximately the same share as J. Division II is gradually and increasingly individualized, and fully human true-life portraits appear in Division III, while Division I has a quasi-mythical nature.

In Phase II, the aim was to let the features arrange themselves into groups of themselves, as it were, without any preconceptions. Five analysis methods were applied, and cluster analysis yielded a most interesting picture of interrelations among the text samples (figures 3.11, 3.12, pp. 136-7). The first cluster comprised samples of Text P. The second cluster comprised most of the N samples outside P. The third cluster encompassed all the H and D samples, regardless of their E or J origin. The authors write explicitly (p. 186): "This adds force . . . to the impression that the Jahwist and the Elohist were each other's alter ego". The other analyses supported this finding.

Lexical richness and lexical concentration were separately examined, yielding results in three different ranges of vocabulary for N,H,D (Fig. 4.3) and J,E,P (Fig. 4.4). Here, too, P is unlike the other groups. After the statistical analysis the authors consider it justified that N,H,D should be distinct. It is doubtful, to their mind, whether J,E,P should be ascribed to three different sources. They summarize this issue as follows: "While solely on these grounds this [documentary] hypothesis cannot be rejected out of hand, such serious doubts regarding its validity have arisen that neither can it any longer be accepted as unreservedly as it has been hitherto." (p. 214)

Rabin in his contribution is not surprised that the sources of discourse have been found so distinctive. Indeed, modern linguistics tends to take into account various extra-linguistic (psycholinguistic and sociolinguistic) factors that yield distinct linguistic patterns in dialogue, narration, etc. Narrative is known to differ from direct speech in various linguistic elements, such as use of pronouns, verb tenses, and demonstratives. It is, then, little wonder that Deity should use an elevated style, full of rhetorical elements on the one hand, and complete and complex sentences on the other, as found in Genesis. He suggests examining the Book according to discourse analysis techniques and the comparison of results.

Talmon warns the reader against the spell of modern technological aids, since another computer-aided study of a part of Genesis by a French scholar has yielded different results. He notes that N predominates all over the text of Genesis, with 53% of Division III, 56% of Division II, and 74% of Division I. H increases from 5% in Division I to 34% in Division II and 47% in Division III. D, however, decreases from 21% in Division I to 10% in Division II and complete absence in Division III. This

internal consistency leads him to suggest that N may have used pre-existing Canaanite or generally known Mesopotamian material. Talmon stresses that the kind of text analysis as applied in this book is really a literary analysis, which fits well with modern trends of Bible scholarship.

Belonging to the linguists' part of the expected reading public (which according to the writers includes "Bible scholars, linguists, statisticians, and probably computer people"), I find this book interesting to study. The starting point of the documentary hypothesis is concisely and clearly summarized. The statistical methods of each phase are clearly described, and are illustrated by tables, figures, and appendices. The findings appear at the end of each statistical analysis, with a brief summary of the results (without the numbers). It may be noted that there is no list of references, and references to the literature appear as footnotes. Appendices are said to be four in number, but only three exist, although "appendices 3 and 4" comprise various sub-sections. There is also a printing mismatch of the number of counted words between the text (p. 192) and table 4.1 (p. 193) and exhibit 4.1 (p. 129).

The results of this study make us hope this is not the last word on the subject (see "Suggestions for further investigation and two examples" pp. 187-189, and "Interim Postscript" pp. 216-217). Subsequent studies, perhaps with more details of the linguistic features proper, are to be awaited.

Judith Rosenhouse

Department of General Studies
The Technion, Haifa, Israel

GETTING COMPUTERS TO TALK LIKE YOU AND ME: DISCOURSE CONTEXT, FOCUS, AND SEMANTICS (AN ATN MODEL)

Rachel Reichman

Cambridge, MA: The MIT Press, 1985, xiii+221 pp.
ISBN 0-262-18118-9; \$20.00

In this optimistically-titled book, Reichman presents a theory of discourse processing that is intended to be the basis for a computational module. The claim is that this would give a full-scale natural language processing system the capacity to generate and interpret extended coherent discourse, rather than just short utterances. The module would be one of the many that make up such a system, although Reichman does not consider what other components would be required, and how they would interact with the discourse module.

The module would keep a dynamically-updated record of the discourse it is engaged in, in such a way as to help it to produce conversational moves that are relevant to the discourse and contain correct use of pronominalization and other discourse phenomena. Reichman provides

a very small part of the algorithm upon which the program to do this might be based, although this is not sufficient to give more than a flavour of what the architecture of that program might be.

The discourse module itself is presented as an ATN grammar which processes detailed specification of context spaces – Reichman's super-sentential unit. The context space is an abstract structure of slots that contains a conversational move as well as a great deal of other information such as what the goal of the move is, what the method of achieving that goal is, who the speaker is, what other context space this space is particularly closely related to, and much more. This specification is interrogated by tests in the ATN as the space is processed, and the results are used to set registers that form the record of the current discourse move. The ATN produces one conversational move specification per pass, using its register values to provide continuity between passes, and hence continuity in the discourse structure it specifies as output.

Reichman claims that her program is written to enable a computer to be a conversational participant, but the examples are simulations of entire discourses between several participants, with the computer playing all the parts itself. Conversation with a human participant is assumed to be possible with "minor" adaptations, but person-machine interaction is not actually addressed in the model presented in the book. It is not clear that the transformations required to get the program to act as a conversational participant would be at all minor, since this would require at least the addition of a "hearer model" of the discourse, and comprehensive repair capabilities for the occasions when speaker and hearer models might disagree. The difficulties in classifying the current conversational move would be greatly magnified if the machine did not have access to a detailed analysis of that move.

Perhaps the most serious problem with Reichman's theory is the fact that she feels she has isolated rules that apply in any sort of discourse, "regardless of its purpose and subject matter, and regardless of the situational context ...". All Reichman's data is composed of conversations between equals, on topics such as politics and personal relationships. I applied the theory to some question/answer data of the type that is very likely to be encountered by just such a system as Reichman is aiming at: I found enough difficulties with context space categories and the choices offered by the ATN to make me very uneasy about the universality of her rules. These are not simply problems caused by the incompleteness of the program, which Reichman admits, but have to do with the decomposition of data into segments for processing, possibly meaning that this operation will have to be changed to deal with certain types of data in ways that would change the character of the context space. This

could have been avoided by an analysis of more information-exchanging rather than argumentative-type data.

Another problem is the ATN itself. Firstly, the analogy drawn between this model and a sentence ATN is misleading, since it is not specified at all clearly what the input or output is here, or how the input would be consumed. Secondly, it seems that the ATN diagram is a sketch of what would take place during processing, omitting many details of the anatomy of the program. The arc tests are labelled but not documented, and the nodes themselves must conceal large portions of code. It therefore seems that the algorithm that would be central to the module is not that shown in the diagram, even though its importance is underlined by the title.

Reichman's outline definitely does not provide enough detail to render implementation trivial. It may be that the difficulty of mapping context space specifications from the surface structure of utterances will be beyond the state of the art for a long time. Another more general problem is that if the division of labour between "boxes" in an NLP system is not well worked-out in advance, it is easy for certain tasks to fall down the cracks between them. There is little attention paid to this in the book.

These considerations apart, as an abstract theory of discourse processing Reichman's work does bring out some important ideas, especially in the way the theory indicates how much of the discourse context is relevant for the processing of a given utterance. What is frustrating is that the work would be so much more useful and easy to evaluate if the mechanism to connect the discourse module to the surface structure of utterances was available, or even talked about. At the moment, applying the framework to the data manually requires so much intuition that the point of details given at higher levels is effectively lost. The incompleteness of the program outline and the inadequacy of presentation of the "ATN" are also a real hindrance.

The final three chapters of the book provide a good overview of the theories that deal with the coherence of discourse, and provide an idea of where Reichman considers her theory to have its peers. Chapters 5 and 9 also include discussion of the work of Grosz (e.g., 1977) and Sidner (1979), whose work is perhaps closest to Reichman's own. However, it is not clear, especially in the case of Sidner, that where Reichman points out theoretical differences between her approach and Sidner's, the differences are either substantial or a significant improvement.

This is the clearest summary of Reichman's work to date, and does make the contents of her thesis (Reichman 1981) much more accessible. I would recommend it to anyone wishing to get to know the research of one of the major names in computational discourse processing, but for those expecting to do either discourse analysis or programming based on the book – approach with caution.

Judy Delin
Centre for Cognitive Science
University of Edinburgh
Edinburg EH8 9LW
Scotland

REFERENCES

- Grosz, B.J. 1977 The Representation and Use of Focus in Dialogue Understanding. Technical Note No. 151, Stanford Research Institute Artificial Intelligence Center, Menlo Park, California.
- Reichman, R. 1981 Plain Speaking: A Theory and Grammar of Spontaneous Discourse. Report No. 4681, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.
- Sidner, C.L. 1979 Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Technical Report No. 537, MIT Artificial Intelligence Laboratory, Cambridge, Massachusetts.

BOOKS RECEIVED

Authors and publishers who wish their books to be considered for review in **Computational Linguistics** should send a copy to the book review editor,

Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, Canada M5S 1A4.

All books received will be listed, but not all can be reviewed.

Readers who wish to review books for the journal should write to the book review editor outlining their qualifications. Obviously, we cannot promise the availability of books in anyone's exact area of interest.

Agosti, Maristella

[Facoltà di Scienze Statistiche, Università di Padova]
Database Design: A Classified and Annotated Bibliography
(The British Computer Society Monographs in Informatics)
Cambridge University Press, 1986, 92 pp.
ISBN 0-521-31123-3 (paper), \$16.95

Andersen, Francis I. and Forbes, A. Dean

[University of Queensland and Hewlett-Packard, resp.]
Spelling in the Hebrew Bible: The Dahood Memorial Lecture
(Biblica et orientalia 41)
Rome: Biblical Institute Press, 1986, xxii+379 pp
ISBN 88-7653-342-7 (paper)

Bennett, Paul A.; Johnson, Rod L.; McNaught, John; Pugh, Jeanette; Sager, J.C.; and Somers, Harold L.

[Centre for Computational Linguistics, University of Manchester Institute for Science and Technology]
Multilingual Aspects of Information Technology
Aldershot: Gower, 1986, ix+146 pp
ISBN 0-566-03513-8 (hardcover), £17.50

Benson, Ian (Editor)

[SRI International]
Intelligent Machinery: Theory and Practice
[Papers presented at a conference organized by SRI International and held in Cambridge, UK, September 1984]
Cambridge University Press, 1986, xi+168 pp
ISBN 0-521-30836-4 (cloth), \$34.50

Chomsky, Noam

Barriers
(Linguistic Inquiry Monographs 13)
Cambridge, MA: The MIT Press, 1986, viii+102 pp.
ISBN 0-262-03118-3 (hardcover)
ISBN 0-262-53067-8 (paper), \$7.95

Demers, Richard A. and Farmer, Ann K.

A Linguistics Workbook
Cambridge, MA: The MIT Press, 1986, xi+275 pp.
ISBN 0-262-54045-2 (paper), \$9.95

Dobbs, James B.

[Lehigh University]
Homophones and Homographs: An American Dictionary
Jefferson, NC: McFarland, 1986, viii+264 pp
ISBN 0-89950-182-6 (hardcover), \$27

Gazdar, Gerald (Editor)

[School of Social Sciences, University of Sussex]
Computational Tools for Doing Linguistics
[Special issue of Linguistics, 23(2), 1985, 185-360]
Berlin: Mouton
ISSN 0024-3949

Grosz, Barbara J.; Sparck Jones, Karen; and Webber, Bonnie Lynn (Editors)

[Harvard University, University of Cambridge, and University of Pennsylvania, resp.]
Readings in Natural Language Processing
Los Altos, CA: Morgan Kaufmann Publishers, 1986,
xv+664 pp.
ISBN 0-934613-11-7 (paper), \$26.95

Hutchins, William John

[University of East Anglia]
Machine Translation: Past, Present, Future.
(Ellis Horwood series in computers and their applications)
Chichester: Ellis Horwood, 1986, 382pp.
ISBN 0-85312-788-3 (hardcover), US\$49.95,
CDN\$111.95

Johnson-Laird, Philip N.

[Medical Research Council, Applied Psychology Unit, Cambridge, England]
Mental Models
(Cognitive Science Series 6)
Cambridge, MA: Harvard University Press, 1983,
xiii+513 pp.
ISBN 0-674-56881-8 (cloth), \$27.50
ISBN 0-674-56882-6 (paper) \$12.95