

ICAME NEWS

Newsletter of the International Computer Archive of Modern English (ICAME). Published by NAVF's EDB-senter for humanistisk forskning, P.O.Box 53, University of Bergen, N-5014 Bergen, Norway. Editor: Dr. Stig Johansson, Department of English, University of Oslo, Norway.

Number 2

March 1979

INTRODUCTORY

Since our first newsletter appeared in March 1978, work has advanced significantly on our three main projects. Brief progress reports will be given below. Two related projects will also be presented: the Syntax Data Project at the University of Gothenburg and the Leuven Drama Corpus. Both projects make material available to other researchers, on conditions specified below.

This issue further contains a bibliography of publications relating to the Brown Corpus. Note also especially the remarks on updating of our mailing list on p. 14.

The next issue of ICAME NEWS is scheduled to appear in September/October 1979. The Editor is grateful for any information or documentation which is relevant to the field of concern of ICAME.

THE BROWN CORPUS

The Bergen versions of the Brown Corpus are now available from the Norwegian Computing Centre for the Humanities, Bergen. These contain upper- and lower-case letters and a minimum of special codes. There are two versions:

Text I: Typographical information is preserved; the same line division is used as in the original version except that words at the end of the line are never divided.

Text II: Typographical information is reduced; the line division is new.

Also available is a microfiche concordance (see *ICAME NEWS* 1). In addition, the concordance is now available on magnetic tape. See the order form at the end of the newsletter.

The tagged version of the Brown Corpus mentioned in *ICAME NEWS* 1 is not yet ready for distribution.

THE LANCASTER-OSLO/BERGEN CORPUS (LOB)

The Corpus is now completed and will be ready for distribution in September/October 1979. For technical information, see the order form at the end of the newsletter.

THE LONDON-LUND CORPUS

The status of the London-Lund Corpus was summarized in the following way in *The Survey of Spoken English: Research Report for 1978* (compiled by Bengt Oreström, Jan Svartvik, and Cecilia Thavenius):

The input and processing of the texts at Lund take place on a video display terminal (Delta 5500) with 96 characters, upper and lower case. The texts are transferred via cassettes (Datum 4002 Universal Recorder) by on-line connection into ASCII-characters in a file on disc at the 1108 UNIVAC computer belonging to Lund University.

By the end of 1978 we had received 87 texts from London, all of which have been edited, put in via the terminal and proofread first time; 70 texts have been proofread second, 41 texts third, and 34 texts fourth time.

A series of check programs is run for spotting logical errors in the input material, such as omitted matching brackets. In addition to the check programs, computer-produced concordances of each text have proved a useful tool for proofreading as well as for research purposes. Concentrating our efforts on text categories 1-3, which represent casual, surreptitiously recorded conversation, we have now produced concordances for each of the 34 texts included in these categories. In the course of 1979 we plan to produce a magnetic computer-tape of this part of the corpus for general distribution. In addition, there will be a printed version of these texts,

A Corpus of Spoken English, to be produced by photosetting technique from the magnetic computer-tape and published in *Lund Studies in English*.

Work on the linguistic analysis of the corpus has also begun. A system for analysing and coding word-classes has been devised and will be implemented as soon as programming facilities allow.

The following books, reports and articles have been produced by members of the team during 1978:

Bengt Oreström, Turn-taking: On Speaker-shift in Face-to-face Conversation. SSE Report.

Bengt Oreström, Avbryt inte!, *Forskning och Framsteg* 3, 28-30.

Bengt Oreström & Cecilia Thavenius, Auditory and Acoustic Analysis: An Experiment. SSE Report.

Randolph Quirk & Jan Svartvik, A Corpus of Modern English. SSE Report; also in *Text-Corpora, Materialien für eine empirische Sprach- und Literaturwissenschaft*. Ed. by H. Bergenholtz & B. Schaedler.

Jan Svartvik & Olof Sager, *Engelsk universitetsgrammatik*. Stockholm: Esselte Studium.

Cecilia Thavenius, A Pilot Study of Referential 'it' in Spoken English. SSE Report.

Three doctoral dissertations based on the corpus are in progress at Lund University: Bengt Oreström on "Interruption and Interrupted Speech"; Anna-Brita Stenström on "Questions and Answers"; Cecilia Thavenius on "Reference in English Conversation".

THE SYNTAX DATA PROJECT

For a number of years Alvar Ellegård (Department of English, University of Gothenburg) and his students have been working on a project which involves grammatical tagging of parts of the Brown Corpus. Results of this work are presented in Ellegård's recently published book on *The Syntactic Structure of English Texts: A Computer-Based Study of Four Kinds of Text in the Brown University Corpus* (Gothenburg Studies in English 43, Göteborg: Acta Universitatis Gothoburgensis, 1978). The four text categories studied are: A (Press: reportage), G (Belles lettres), J (Science), and N (Adventure and western fiction). Sixteen texts from each of these categories, in all about an eighth of the whole corpus, were selected and subjected to a detailed, largely traditional, surface-structure analysis on three levels:

- 1) clause structure of sentences;
- 2) constituent structure of clauses; and
- 3) word class of individual words.

Using an instruction booklet produced by the leader of the project undergraduate students analyzed texts as part of their English course. The results of the analysis were checked by the project leader and/or discussed at seminars. The analyzed text was punched with each word forming a record structured in the following way:

<u>Description</u>	<u>Number of characters</u>	<u>Column(s)</u>
Number of the sentence within the text	6	1-6
Word number within the text	6	7-12
Word	22	13-34
Clause-initial marker	1	35
Brown Corpus identification number	7	36-42
Clause marker	16	43-58
Sequence number with discontinuous clauses	1	59
End of sentence	1	60
Part of the sentence	10	61-70
Word class	5	71-75

A sample printout of some sentences of the analyzed text is given in Table 1. As the system is quite complex, it is only possible in this context to give a rough indication of how it works.

On the clause-structure level there is a classification according to clause types (1st letter symbol: Z = non-subordinate; P = subordinate, non-relative; R = subordinate, relative), clause modality (2nd letter symbol: D = declarative, Q = interrogative, J = imperative, E = exclamatory) and kind of predicate verb (3rd letter symbol: F = finite, J = infinitive, G = gerundive, P = participial, L = verbless). ZDF in Table 1 therefore means a non-subordinate, declarative, finite clause, and PDJ stands for a subordinate, non-relative, declarative, infinitive clause. On the clause level there is also an indication of the degree of embedding (numerical symbols following the letter codes) and of the function of embedded clauses within the superordinate structure (constituent symbol followed by a bracket). For examples, see Table 1.

Table 1 The Syntax Data Project: sample output

000263000944DAWN	*N020740ZDF1	S	N
000263000945WOULD	N020740ZDF1	V	MD
000263000946COME	N020740ZDF1	V1	V
000263000947SOON	N020740ZDF1	C	A
000263000948AND	*N020750ZDF2	Y	Y
000263000949THE	N020750ZDF2	S1	T
000263000950NIGHT	N020750ZDF2	S	N
000263000951WAS	N020750ZDF2	V	BD
000263000952AT	N020750ZDF2	PP	P
000263000953ITS	N020750ZDF2	P1	RX
000263000954COLDEST	N020750ZDF2	4P	JT
000266001056GAVIN	*N020820ZDF1	S	C
000266001057SIGNED	N020820ZDF1	V	VD
000266001058BITTERLY	N020820ZDF1	4D	A
000267001059IN	*N020820ZDF1	AP	P
000267001060THAT	N020820ZDF1	A1	E
000267001061INERT	N020820ZDF1	A2	J
000267001062LANDSCAPE	N020830ZDF1	A	N
000267001063THE	N020830ZDF1	S1	T
000267001064CARAVAN	N020830ZDF1	S	N
000267001065OF	N020830ZDF1	S2P	P
000267001066HIS	N020830ZDF1	S21	RX
000267001067DESIRES	N020830ZDF1	S2	NS
000267001068PASSED	N020830ZDF1	V	VD
000267001069BEFORE	N020830ZDF1	BP	P
000267001070HIS	N020830ZDF1	B1	RX
000267001071MIND	N020830ZDF1	4B	N
000269001088IN	*N020840ZDF1	EP	P
000269001089VAIN	N020850ZDF1	E	N
000269001090HIS	N020850ZDF1	S1	RX
000269001091MIND	N020850ZDF1	S	N
000269001092GROPED	N020850ZDF1	V	VD
000269001093/PDJ11	N020850ZDF1	E/	
000269001094TO	*N020850PDJ11	VP	U
000269001095REASSEMBLE	N020850PDJ11	V	V
000269001096THE	N020850PDJ11	O1	T
000269001097BONES	N020850PDJ11	O	RS
000269001098OF	N020850PDJ11	O2P	P
000269001099THE	N020850PDJ11	O21	T
000269001100RELATIONSHIPS	N020850PDJ11	O2	NS
000269001101/RDF111	N020850PDJ11	O22/	
000269001102	*N020850RDF111	10-	
000269001103HE	N020850RDF111	S	R
000269001104HAD	N020850RDF111	V	HD
000269001105SOUGHT	N020860RDF111	V1	VN
000269001106SO	N020860RDF111	D1	A
000269001107DESPERATELY	N020860RDF111	D	A
000269001108BUT	*N020860ZDF2	AY	Y
000269001109THEY	N020860ZDF2	S	RS
000269001110WOULD	N020860ZDF2	V	MD
000269001111NOT	N020860ZDF2	M	G
000269001112COME	N020860ZDF2	V1	V
000269001113TO	N020860ZDF2	BP	P
000269001114LIFE	N020860ZDF2	4B	N
000272001171THEN	*N020900ZDF1	C	A
000272001172AT	N020900ZDF1	CP	P
000272001173LAST	N020900ZDF1	C	JT
000272001174THE	N020900ZDF1	S1	T
000272001175DANKLESS	N020910ZDF1	S	N
000272001176BEGAN	N020910ZDF1	V	VD
000272001177/PDJ11	N020910ZDF1	O/	
000272001178TO	*N020910PDJ11	VP	U
000272001179DISSOLVE	N020910PDJ11	4V	V

On the constituent-structure level the following main elements are recognized: S = subject, O = direct object, non-prepositional, J = indirect object, non-prepositional, K = prepositional object, V = primary predicative verb. Other constituents are adverbs of place (A), direction (D), time (C), modality (M), manner and degree (D), other adverbs (E), agent of passives (G), existential there (F), predicate complement of subject (P), predicate complement of object (Q), addressed person (T), interjection (I), pure co-ordinating conjunction (Y), pure subordinating conjunction (Z), adverbial co-ordinating conjunction (AY), adverbial subordinating conjunction (AZ). For examples of some of these categories, see Table 1.

Constituents of noun phrases, verbal groups, adjective phrases, prepositional phrases, etc. are indicated by giving the modifier (prepositions are looked upon as modifiers of the following noun, main verbs as modifiers of the preceding auxiliaries) the same constituent symbol as the head, plus an indicator of the number of the modifier (in the case of prepositions P is used rather than a numeral). As examples, consider at its coldest or the caravan of his desires in Table 1.

The word-class level contains separate markers for major form classes (noun, verb, etc.), function words (prepositions, conjunctions, etc.), and inflectional categories. As the word-class marking is relatively straightforward, it will not be further dealt with here.

Ellegård's detailed tagging system makes it possible to have access to very narrowly defined structures, such as that-clauses functioning as subject or object, fronted objects or predicate complements, predicative superlative adjectives, etc. Though the marking is mainly based on surface structure, an extraposed clause (as in it is difficult to see him) is marked as subjective. Structures such as I wanted him to go and I asked him to go are further marked differently. The system allows for unambiguous marking of discontinuous constituents. Nevertheless, there are a number of problems and inadequacies. As Ellegård himself points out (p. 93), "no distinction is made between a fine little house and a large expensive house, though it may be held that in the first instance we refer to a little house that is

fine, and in the second to a house that is both large and expensive". Zero that-clauses are difficult to identify unambiguously (whereas zero relative clauses form no problem). There are also numerous problems in the tagging of sentence constituents, as is to be expected. A selection of "problematic sentences" is given in an appendix of the book.

In his book Ellegård presents the results of a large number of frequency counts relating to the three levels of analysis: number of clauses per sentence, types of clauses, constituent structure of clauses, word-class distribution, phrase length, depth of modification, etc. These results, though they are often not very surprising, add significantly to our knowledge of English grammar in use.

In the concluding section of the book Ellegård stresses that "the chief usefulness of our work lies in the possibility it offers to use the computer for quite sophisticated searches for specific grammatical and stylistic features ..." (p. 79) and he points out that very much more information can be extracted from his material than is offered in the present work. We hope that he will return later with further investigations. Meanwhile, we note with gratefulness that Ellegård invites other researchers to make use of his material (p. 7): "Copies of the tape are available for research purposes at cost price (at present--1978--c. 400 Swedish kronor, or 100 US dollars). Both the texts and the analysis are available in microfiche on request (13 fiches, c. 1 dollar per fiche)." Specification: record length = 75, block length = 7500, code = EBCDIC, density = 1600 bpi, labelling: IBM standard labels.

Those who are interested in obtaining Ellegård's material should write to: Professor Alvar Ellegård, Department of English, University of Gothenburg, Lundgrensgatan 7, S-412 56 Göteborg, Sweden. The book dealt with above is available from: ACTA UNIVERSITATIS GOTHOBURGENSIS, Box 5096, S-402 22 Göteborg 5, Sweden.

THE LEUVEN DRAMA CORPUS

The corpora of modern English texts dealt with previously do not include any dramas. It is therefore important to draw attention to the Leuven Drama Corpus, which has been compiled at the Institute of Applied Linguistics of the Catholic University of Leuven under the direction of Professor L. K. Engels and Dr. Dirk Geens. The corpus contains 62 British

English plays first published during 1966-1972. The plays were "selected in such a way that the sample is homogeneous for syntactic research and representative of the literary genre that provides the material" (Geens 1978: 46). Among the principles used in selecting the material could be mentioned: (1) the sources had to be original (untranslated) plays produced by British English authors; (2) they had to consist of more than one act; (3) they were to contain a minimum of dialect, archaic or lyrical passages; (4) the spelling must not differ from traditional norms; (5) only one play per author could be included. In all, the corpus contains about a million words, i.e. it is comparable in size to the Brown Corpus.

The text of the plays was punched in capitals and with a minimum of editing. Each sentence was started on a new record. Among editing conventions could be mentioned that subject+auxiliary contractions were dissolved (e.g. he'll becomes he 'll) and that a special end-of-sentence symbol was inserted.

The Leuven Drama Corpus has been studied on various levels, especially by Dirk Geens (see below): word frequencies, word and sentence length, grammatical analysis. However, the material could doubtless be used in further investigations. It is therefore of great interest to note that copies of the Corpus, or portions thereof, are available from the compilers on request. Specification: record length = 60, block length = 6000, density = 1600 bpi, code = EBCDIC, labelling: IBM standard labels. Price: 10,000 BF for the whole corpus. Those who wish to order the material should write to: Professor L. K. Engels, Instituut voor Toegepaste Linguistiek, Tiense Vest 27, 3200 Kessel-LO, Belgium OR Dr. Dirk Geens, Afdeling Toegepaste Linguistiek, Blijde-Inkomststr. 21, B-3000 Leuven, Belgium. Some publications relating to the Leuven Drama Corpus are:

Dé Kock, J., Geens, D., Por una estilística lingüística y cuantitativa con la ayuda del ordenador. Madrid, Revista de la Universidad Complutense, Vol. XXV, Nr. 102, 1976, 113-129.

De Kock, J., Geens, D., Pour une stylistique linguistique et quantitative avec l'aide de l'ordinateur. Paris, Linguistique Automatique et Langues Romanes (ed. J. De Kock, C. Boitet), 1977, 198-215.

Geens, D., A Frequency-List of Sentence Structures: Preliminary Considerations. Leuven, Review of the ITL, Nr. 21, 1973, 39-55.

Geens, D., A Venture in Computational Analysis of English Syntax. Leuven, Preprints of the Department of Linguistics, Nr. 19, 1974.

Geens, D., Analysis of Present-Day English Theatrical Language. Leuven, Progress in Applied Linguistics, Nr. 6 (Progress Report) 1975.

Geens, Dirk, 1978. On measurement of lexical differences by means of frequency. In *Glottometrika* 1, ed. by G. Altmann, pp. 46-72. Bochum: Studienverlag Dr. N. Brockmeyer.

BROWN CORPUS BIBLIOGRAPHY (BY W. NELSON FRANCIS)

One of the aims of ICAME is to distribute information on corpus research. The list below, which has been compiled by W. Nelson Francis (with some additions made by the Editor), specifies works "using or referring to the Brown Corpus". In some cases brief notes on individual works are added within parentheses.

Baron, Naomi S. 1977. Language acquisition and historical change. North-Holland Linguistic Series 36. Amsterdam, New York, and Oxford: North-Holland Publishing Co.

(Uses Brown Corpus *make, get, have* as control in study of acquisition of periphrastic causatives.)

Britton, Bruce K. 1978 (forthcoming). Lexical ambiguity of words used in English text.

Burton, Dolores M. 1968. *Respice finem* and the *tantum quantum*: an essay review of computational studies for 1967-68. *CHum* 3.41-48. (Review of Kučera and Francis 1967, pp. 44-45.)

Card, William, and Virginia McDavid 1966. English words of very high frequency. *College English* 27.596-604.

Creswell, Thomas J. 1975. Usage in dictionaries and dictionaries of usage. Publication of the American Dialect Society, Nos. 63-64. University, Alabama: Alabama University Press.
(pp. 135-40 based on Meyers 1972)

Dubois, Betty Lou 1972. Meaning and distribution of the perfect in present-day American English prose. Unpublished Ph.D. dissertation, University of New Mexico. DAI 33/12-A, pp. 6892f.

Dubois, Betty Lou, and Isabel M. Crouch 1978. *Man and its compounds* in recent prefeminist American English published prose. Mimeo: New Mexico State University, Las Cruces, N.M.

Ehrman, Madeline E. 1966. The meanings of the modals in present-day American English. *Janua linguarum, series practica* 45. The Hague: Mouton.

Ellegård, Alvar 1978. The syntactic structure of English texts: A computer-based study of four kinds of text in the Brown University Corpus. *Gothenburg Studies in English* 43. Göteborg: Acta Universitatis Gothoburgensis. (Presents results of a detailed grammatical analysis of about an eighth of the Corpus; see p. 3 of this issue)

Enkvist, Nils Erik 1973. Linguistic stylistics. *Janua linguarum, series critica* 5. The Hague: Mouton.

- Filipović, Rudolf 1969- The Yugoslav Serbo-Croatian - English contrastive project. A. Reports, B. Studies, C. Pedagogical materials. Zagreb: Institute of Linguistics; and Washington: Center for Applied Linguistics. (Uses $\frac{1}{2}$ of Brown Corpus translated into Serbo-Croatian for contrastive study with original.)
- Francis, W. Nelson 1964, 1971. Manual of information to accompany A Standard Sample of Present-Day American English, for Use with Digital Computers. Revised Ed. Providence: Department of Linguistics, Brown University.
- 1964a. A standard sample of present-day English for use with digital computers. Report to U.S. Office of Education on Cooperative Research Project No. E-007. Providence, Brown University.
- 1965. A standard corpus of edited present-day American English for computer use. In *Literary Data Processing Conference Proceedings, September 9, 10, 11, 1964*, ed. Jess B. Bessinger, Jr., Stephen W. Parrish, and Harry F. Arader. Armonk, N.Y.: IBM Corporation, pp. 79-89. A slightly revised version in *College English* 26.267-73.
- 1967. The Brown University Standard Corpus of English: some implications for TESOL. In *On teaching English to speakers of other languages*, ed. Betty Wallace Robinett, pp. 131-5. Washinton: TESOL.
- 1975. Problems in assembling, describing, and computerizing corpora. In *Papers in Southwest English 1, Research techniques and prospects*, ed. Betty Lou Dubois and Bates Hoffer, pp. 15-38. San Antonio: Trinity University.
- 1978. Problems of assembling and computerizing large corpora. In *Text-Corpora: Materialien für eine empirische Sprach- und Literaturwissenschaft*, ed. by H. Bergenholtz and B. Schaefer. Kronberg: Scriptor Verlag. (Revised version of preceding.)
- Forthcoming. A tagged corpus--problems and prospects. (Written Dec. 1977: to be published in July 1980.)
- Geens, Dirk 1978. On measurement of lexical differences by means of frequency. In *Glottometrika 1*, ed. by G. Altmann, pp. 46-72. Bochum: Studienverlag Dr. N. Brockmeyer. (Compares word frequencies in the Brown Corpus and the Louvain Drama Corpus)
- Greene, Barbara B., and Gerald M. Rubin 1971. Automatic grammatical tagging of English. Providence: Department of Linguistics, Brown University.
- Haskel, Peggy I. 1971. Collocations as a measure of stylistic variety. In *The computer in literary and linguistic research*, ed. R. A. Wisbey, pp. 159-68. Cambridge University Press.
- Hauge, Jostein and Knut Hofland 1978. Microfiche version of the Brown University Corpus of Present-Day American English. Bergen: NAVF's EDB-Senter for Humanistisk Forskning. (Text in upper- and lower-case; KWIC concordance.)
- Hermerén, Lars 1978. On modality in English: A study of the semantics of the modals. Lund Studies in English 53. Lund: CWK Gleerup.

- Householder, Fred W. 1971. Linguistic speculations. Cambridge University Press. (See Chapter 13, "The primacy of writing.")
- Johansson, Stig 1978. Some aspects of the vocabulary of learned and scientific English. Gothenburg Studies in English 42. Göteborg: Acta Universitatis Gothoburgensis. (Studies word frequencies in category J of the Brown Corpus.)
- Johansson, Stig, Leech, Geoffrey and Helen Goodluck 1978. Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use With Digital Computers. Oslo: Department of English, University of Oslo.
- Kelly, Edward, and Philip Stone 1975. Computer recognition of English word senses. North-Holland Linguistic Series 13. Amsterdam and Oxford: North-Holland Publishing Co. (On General Inquirer disambiguating routines. Used Brown Corpus for control: see chapter III.)
- Kučera, Henry 1968. Some quantitative lexical analyses of Russian, Czech, and English. In *American Contributions to the Sixth International Congress of Slavists*, I, ed. by Henry Kučera, pp. 1-44. The Hague: Mouton.
- Kučera, Henry, and W. Nelson Francis 1967. Computational analysis of present-day American English. Providence: Brown University Press. (Contains alphabetical and rank-ordered frequency lists, also several studies of word distribution and sentence length in the Corpus.)
- Leech, Geoffrey, and Rosemary Leonard 1974. A computer corpus of British English. Hamburg: Hamburger Phonetische Beiträge 13, pp. 41-57. (Describes the Lancaster Corpus, a parallel to the Brown Corpus.)
- McKinnon, Alistair 1975. Computational analysis of Kierkegaard's *Samlede Verker*. Leiden: E. J. Brill. (Modeled on Kučera-Francis 1967.)
- Meyers, Walter E. 1972. A study of usage based on an examination of the Brown Corpus. In *College Composition and Communication* 23.155-69.
- Monroe, George K. 1965. Phonemic transcription of graphic post-base affixes in English: a computer problem. Unpublished Ph.D. dissertation in Linguistics. DA 26/08, p. 464.
- Morris, William, ed. 1969. The American Heritage dictionary of the English language. New York: American Heritage and Houghton Mifflin. (See "Computers in language analysis and in lexicography," by Henry Kučera, pp. xxviii-xl.)
- Pearson, Charles 1978. Quantitative investigations into the type-token relation for symbolic rhemes. Preliminary version. Georgia Institute of Technology.
- Recktenwald, Robert P. The English progressive: semantics and history. Unpublished Ph.D. dissertation, Brown University. (Used Brown Corpus for 20th century material.)
- Ringbom, Håkan 1973. George Orwell as essayist: A stylistic study. Acta Academiae Aboensis, Ser. A, Humaniora, 44:2. Åbo: Åbo Akademi. (Uses category G of the Brown Corpus as a norm in a study of Orwell's essays.)
- 1975. The style of Orwell's appendix to *Animal Farm*. In *Style and Text, studies presented to Nils Erik Enkvist*. Stockholm: Språkförlaget Skriptor AB, pp. 243-9.

- Schaefer, Burkhard 1976. Maschinenlesbare Textkorpora des deutschen und des englischen. *Deutsche Sprache* 4.356-70.
- Sedelow, Sally Y., and Walter A. Sedelow, Jr. 1967. Stylistic analysis. In *Automated language processing*, ed. H. Borko, pp. 181-213. New York: Wiley.
- Sherman, D. 1977. A computer archive of language materials. In *Computing in the Humanities: Proceedings of the Third International Conference on Computing in the Humanities*, ed. S. Lusignan and J. S. North, p. 283. Waterloo, Ont.: University of Waterloo Press.
- Sinclair, J. McH. 1970. English lexical studies. Final report to OSTI on Project C/LP/08 for period January 1967-September 1969. Birmingham: Department of English, The University.
- Smith, Donald A. 1971. An automatic parsing procedure for simple noun- and verb-phrases. Brown University M.A. thesis in Linguistics.
- Smith, Raoul N. 1968. A quantitative study of Russian vocabulary. Unpublished Brown U. Ph.D. dissertation in Linguistics. DAI 30/01-A, p. 307. ("Much of the organization of the Russian Corpus owes its form to" the Brown Corpus.)
- 1973. Probabilistic performance models of language. *Janua linguarum, series minor*, 150. The Hague: Mouton.
- Stanford University Department of Linguistics n.d. Stanford University computer archive of language materials (CALM). Mimeo.
- Sundby, Bertil 1970. Front-shifted ING and ED groups in present-day English. Lund: C.W.K. Gleerup.
- 1976. The *-ing* form revisited. *Lingua* 40.321-30.
- Svartvik, Jan 1968. Plotting divided usage with *dare* and *need*. *Studia Neophilologica* 40.130-40.
- Tanaka, Harumi 1971. A statistical study on selectional features of transitive verbs in present-day American English. Unpublished Ph.D. dissertation, Brown University. DAI 32/10-A, p. 5769.
- Typographic Council for Spelling Reform (Edward Rondthaler and Edward J. Lias) n.d. Untitled pamphlet on Soundspel, a simplified spelling system by computer.
- Warren, Beatrice 1978. Semantic patterns of noun-noun compounds. *Göteborg Studies in English* 41. Göteborg: Acta Universitatis Gothoburgensis. (Studies compounding relationships in 4,500 noun+noun compounds collected from the Brown Corpus.)
- Zettersten, Arne 1968. Current computing activity in Scandinavia relating to language and literature research. *CHum* 3.53-60.
- 1969. A statistical study of the graphic system of present-day American English. Lund: Studentlitteratur.
- 1969. A word-frequency list of scientific English. Lund: Studentlitteratur. (Contains a rank list of the words of category J of the Brown Corpus.)

ADDITIONS TO THE BROWN CORPUS BIBLIOGRAPHY

We are aware that the list given above may be incomplete and therefore invite our readers to supplement the bibliography. If you are familiar with any publication/s/ not included, please fill in the form below.

Note that we are also interested in major unpublished reports and, especially work in progress. Later issues of our newsletter will contain additions to the Brown Corpus Bibliography already given.

 (Return to: Stig Johansson, Department of English, University of Oslo,
 P.O.Box 1003, Blindern, Oslo 3, Norway)

Published books/articles relating to the Brown Corpus:

Unpublished reports:

Work in progress:

Name and address:

CONFERENCE ON DATA BASES IN THE HUMANITIES

There will be a conference on data bases in the humanities and the social sciences at Dartmouth College, Hanover, N.H. 03755, 23-24 August, 1979. The conference is sponsored by the IFIP Task Group for Humanities and Social Sciences, Queens College, City University of New York, and follows immediately after ICCH/4.

SYMPOSIUM ON GRAMMATICAL TAGGING OF TEXT CORPORA

A symposium on "Grammatical Tagging of English Text Corpora in Machine-Readable Form" will be held at the Norwegian Computing Centre for the Humanities, Bergen, March 29-30, 1979. Preliminary programme:

Randolph Quirk (University College London), "On the Place of a Corpus in English Language Research"

W. Nelson Francis (Brown University), "A Tagged Corpus: Problems and Prospects"

Jan Svartvik (University of Lund), "Tagging Spoken English"

Jan Aarts (Katholieke Universiteit, Nijmegen), "Grammatical Tagging in the Dutch Computer Corpus Pilot Project"

Alvar Ellegård (University of Gothenburg), "On the Tagging of Sentence Constituents"

Rudolf Filipović (University of Zagreb), "The Grammatical Tagging of the 'Zagreb Version' of the Brown Corpus"

Mamata Nakra (McGill University), "Grammatical Tagging of Journalistic Prose"

Claus Faerch (University of Copenhagen), "Word Class Tagging of Learners' Language"

Dirk Geens (Katholieke Universiteit, Leuven) "Grammatical Analysis of a Present-Day English Drama Corpus"

Henry Kučera (Brown University) "The Frequency of Grammatical Classes in the Brown Corpus"

The next issue of *ICAME NEWS* will contain a report from the symposium.

UPDATING OF THE MAILING LIST

The mailing list for *ICAME NEWS* is now so large that it is essential to eliminate unnecessary entries. If you wish to receive the newsletter in the future, fill in the form below.

(Return to: Stig Johansson, Department of English, P.O.Box 1003, Blindern, Oslo 3, Norway)

I/We wish to receive *ICAME NEWS*

Name/Institution:

Address:

ORDER FORM

The procedure for ordering material has now been changed. To obtain material you must enclose a cheque (bank draft, cashier's cheque) with your order made out in Norwegian currency to: The Norwegian Computing Centre for the Humanities, Bergen, Norway.

(Return to: The Norwegian Computing Centre for the Humanities,
P.O.Box 53, University of Bergen, 5014 Bergen, Norway)

Indicate in the table below what you wish to order (prices are given in Norwegian kroner):

Magnetic tapes	9-track, 1600 FPI		7-track, 800 FPI		7-track, 556 FPI	
	Number of tapes	Price	Number of tapes	Price	Number of tapes	Price
Brown: Text I	1 1200 ft.	150	1 2400 ft.	275	1 2400 ft.	275
Brown: Text II	1 1200 ft.	175	1 2400 ft.	300	1 2400 ft. 1 1200 ft.	300
Brown: Texts I+II	1 1200 ft.	200	1 2400 ft. 1 1200 ft.	525	2 2400 ft.	550
Brown: KWIC concordance	4 2400 ft.	1100	11 2400 ft.	3375	15 2400 ft.	3900
LOB: Text	1 1200 ft.	150	1 2400 ft.	250	1 2400 ft.	250
LOB: KWIC concordance	5 2400 ft.	1250	12 2400 ft.	3650	16 2400 ft.	4100

- ☐ Brown: KWIC concordance (microfiche) Price: 350
☐ LOB: KWIC concordance (microfiche) Price: 350

Please, add for postage and handling:

for each 1200 ft. tape 25 Norwegian kroner
for each 2400 ft. tape 35 " "
for each microfiche set 10 " " (overseas air mail: 20)

Packages will normally be sent by surface mail. Note that the LOB (Lancaster Oslo/Bergen Corpus) material will not be ready for distribution until September/October 1979.

Name/Institution: _____

Address: _____

We understand that the material listed above is for research purposes only and agree not to distribute the material or reproduce any part of it for any other purpose than scholarly research. (To be signed by a responsible official of the institution making the order.)

Date _____ Signed _____

Print name _____

Official Position _____