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 3 October 1979

CONTENTS

The present newsletter contains reports from the three main ICAME projects. The recently completed tagged version of the Brown Corpus is briefly presented, and the LOB Corpus and the material now available from the London-Lund Corpus are described. In addition, this number includes a report from the symposium on grammatical tagging mentioned in *ICAME NEWS* 2 and information on machine-readable texts at the computer centres in Oxford and Cambridge.

THE BROWN CORPUS

The Bergen versions of the Brown Corpus are available, as announced in the preceding number. Also available is a complete concordance on microfiche and magnetic tape. See the order form at the end of the newsletter.

The tagged version of the Brown Corpus which has been referred to previously has now been completed at Brown University under the direction of W. Nelson Francis and Henry Kučera. Computer tapes including the text and grammatical annotation can now be ordered. The following is an excerpt from the announcement about the tagged corpus made by the compilers.

The grammatical analysis of the Corpus is based on a taxonomy of 81 grammatical classes or "tags". These classes represent an expanded and refined system of word-classes, supplemented by major morphological information (e.g. singular vs. plural, tense, etc.) as well as some syntactic information (e.g. subordinate vs. coordinate conjunctions). Each of the one million words of the Corpus has been annotated with its proper "tag". All lexical items that are grammatically ambiguous have been disambiguated. A manual, which contains detailed information about the textual selections in the Corpus, a list and explanation of the grammatical taxonomy, as well as information about the formating of the data on computer-processible magnetic tapes, is also available.

The grammatically annotated Corpus text encompasses two 2400' magnetic tapes, available in a 9-channel format (7-channel format by special arrangement) and the usual densities. We are able to provide either ASCII or EBCDIC coding. The cost of both tapes, which includes the reels, the manual and surface shipping (with air freight additional), is $250.- for universities and other non-profit institutions, and $350.- for all others.

The grammatically analyzed tapes are protected by copyright and no reproduction of them can be made without the written permission of the copyright holders, W.N. Francis or Henry Kučera. However, the tapes can be utilized for the usual research purposes, such as retrieval of grammatical information or automatic parsing.

For those users who are interested in specific information, for example in the occurrences of particular lexical items, word-classes or grammatical forms, we can provide (either in printed form or on computer tape) a KWIC (Key-Word-in-Context) concordance of such items. The desired items can be specified either lexically or grammatically, or as a combination of both of these parameters. The cost of such concordances will depend on the frequency of the items to be retrieved as well as on the

complexity of the specifications for their retrieval. Users interested in this type of information should contact us for an estimate at the address given below.

In the near future, we also expect to have available a lemmatized frequency list of the Brown Corpus. This frequency list will group all inflected forms of a base form or a "lemma" under one heading, with the appropriate tags for each form, and the total frequency for each lemma as well as for each subentry given. We also plan to provide statistical information about the occurrence of grammatical forms in the Corpus as a whole as well as in each genre of the Corpus.

A special order form for the grammatically annotated tapes is included at the end of this newsletter. Further information on the tagged Brown Corpus can be obtained from Henry Kučera, TEXT RESEARCH, 196 Bowen Street, Providence, R.I. 02906, U.S.A.

THE LOB CORPUS

The Lancaster-Oslo/Bergen Corpus (LOB) is now ready for distribution, as announced in *ICAME NEWS* 2. The Corpus contains about a million words of printed British English texts published in 1961, representing 500 samples of about 2,000 words distributed over 15 text categories. It is intended to be a British counterpart of the Brown Corpus and is therefore, as far as possible, comparable with its American equivalent. The size, year of publication of the texts, sampling principles, etc. are identical. Table 1 summarizes the basic composition of the two corpora. A full description of the LOB Corpus, including a detailed comparison with the Brown Corpus, is given in the *Manual* referred to below.

Research on the material has already started. Word frequency lists and detailed comparisons with word frequencies in the Brown Corpus will be published shortly in a book by Stig Johansson and Knut Hofland. Investigations of modal verbs, the genitive, coordination, and existential <u>there</u> are in progress as well as several minor studies of other aspects.

Table 1 The basic composition of the British and American corpora

Text categories		Number of texts in each category	
		American corpus	British corpus
A	Press: reportage	44	44
B	Press: editorial	27	27
C	Press: reviews	17	17
D	Religion	17	17
E	Skills, trades, and hobbies	36	38
F	Popular lore	48	44
G	Belles lettres, biography, essays	75	77
H	Miscellaneous (government documents, foundation reports, industry reports, college catalogue, industry house organ)	30	30
J	Learned and scientific writings	80	80
K	General fiction	29	29
L	Mystery and detective fiction	24	24
M	Science fiction	6	6
N	Adventure and western fiction	29	29
P	Romance and love story	29	29
R	Humour	9	9
Total		500	500

The LOB text is in upper- and lower-case and includes some "interpretive" coding designed to prepare the material for linguistic analysis: markers of abbreviations and non-English material, separate symbols for the beginning and the end of quotations, headline codes, and sentence-initial markers, etc. Work on grammatical tagging of the corpus is in the planning stage.

The LOB text can be ordered through ICAME as well as a complete concordance on microfiche and magnetic tape. The concordance contains both an exhaustive list of the occurrences of each word in context and frequency information. To facilitate a comparison, the LOB concordance also includes the words and frequency figures for the Brown Corpus. At the end of each entry, there is information on the total frequency of the word in the two corpora as well as detailed information on the distribution of the word in the 15 text categories. For an example, see Table 2.

For technical information on the material available, see the order form at the end of the newsletter. A *Manual* which describes the contents of the Corpus, the sampling principles, coding system, etc., will be distributed together with copies of the magnetic tape of the LOB text:

Johansson, S., in collaboration with Leech, G.N., and Goodluck, H., *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers.* Department of English, University of Oslo, 1978.

THE LONDON-LUND CORPUS

The Survey of English Usage, under the direction of Randolph Quirk, University College London, and the Survey of Spoken English, under the direction of Jan Svartvik, University of Lund, have produced a corpus of spontaneous conversation in orthographic transcription with prosodic analysis. It is now made available both in print and in machine-readable form on computer tape, which will make it possible for interested scholars in any part of the world to have convenient access to this valuable material

Table 2 An extract from the LOB concordance

L16	83	2	value of hypnotic treatment lies in the increased	suggestibility of the patient and also what we call
G55	54	1	listener absorbed. Sheer fatigue might increase	suggestibility.) There was, in the Chinese manner, much
			WORDS=2, G=1, L=1, SAMPLES=2, G=1, L=1, CAT=2	
BROWN			WORDS=3, J=3, SAMPLES=1, J=1, CAT=1	
A17	54	9	public, which has shown itself infinitely	suggestible, knowing nothing between uncritical enthusiasm
F30	62	3	in his struggle for power with his brother, by	suggesting a deceitful way of taking possession of Leon. A
G27	82	1	could well have dispensed with, and even humorously	suggesting a special service of intercession at ØØSt. Paul's
E10	170	8	The left arm could be making one of many gestures	suggesting excitement, and to link up with it the girl's
G69	158	7	Such magical, mysterious, awe-inspiring, divinity-	suggesting facts have included wholly outer phenomena like
B21	141	6	paper a description of Budapest in the early evening	suggesting it was a dark and depressing city. I can
G67	119	7	were capable of altering theoretical assumptions, by	suggesting new ones. Put in this way, Weber's procedure
C12	199	6	his own language, which seems hammered out, a medium	suggesting sheets of gold leaf. *<SEARCHER FOR ATLANTIS*>
A04	185	1	benefits paid during strikes and lock-outs.	Suggesting that a total æØT.U.C.á membership of eight million
D06	13	11	but to a child they are very real. Not that I am	suggesting that children should be molly-coddled- they must
C13	128	1	yet he seems to want to dignify it- oddly enough, by	suggesting that it is a poem of wit which, like Donne's, is
J24	139	6	confirmed. It seems that the authors are correct in	suggesting that it is only rarely that average to bright
K06	9	1	probably thought a convulsive shudder. 'If you're	suggesting that Martin isn't old enough...' she said in a new
P12	97	4	bear to sit still. Even Bridget was no longer	suggesting that the girl had been caught out in an escapade;
B23	67	7	Clearly our two Democratic Socialists are	suggesting that the Labour Party should give up its heritage
C12	129	6	managed adroitly the humorous atrocity story	suggesting that the ministerial troops should castrate the
G47	55	5	is there. This irritation is sometimes relieved by	suggesting that the painting be viewed as a piece of wall
B10	15	7	in the accounts. ØØMr. Aucott is not, I hope,	suggesting that the standards of honesty in British companies
D03	220	3	It carries the idea of reward or profit. Qoheleth is	suggesting that there is a gain from human experience. He has
F08	68	2	and asking him to accompany her, perhaps	suggesting that to make it entirely "her" evening, he allows
G73	101	10	seriously; it seemed to put responsibility on me for	suggesting what the difficulties might be and how they might
			WORDS=20, A=1, B=3, C=3, D=2, E=1, F=2, G=5,	J=1, K=1, P=1, SAMPLES=19, A=1, B=3, C=2,
			D=2, E=1, F=2, G=5, J=1, K=1, P=1, CAT=10	
BROWN			WORDS=13, B=1, D=1, E=1, F=1, G=2, H=2, J=3,	K=1, L=1, SAMPLES=13, B=1, D=1, E=1, F=1, G=2,
			H=2, J=3, K=1, L=1, CAT=9	
F40	151	11	from the plant is serious enough to merit almost any	suggestion aimed at controlling it, and this one, put forward
P07	139	12	for an early marriage. She was not adverse to the	suggestion, but he had to use a deal of pressure before she
J25	87	1	due to apparent size under limited conditions. This	suggestion could be tested experimentally by keeping one
J02	174	8	a warmer climate than today. This was the first	suggestion for an investigation into palaeoclimatology, a
F40	146	10	against them. Last week it was reported that the	suggestion had been put forward to use the coypu to combat
J38	196	2	be the case were the children not streamed. The	suggestion here is that since older children of a year group
G13	156	2	been suggested that he should go but he "evaded the	suggestion." "I travel badly," he tells us, "and I speak
J25	19	3	in fact be produced in this way. Nevertheless, the	suggestion is an interesting one and could be followed up by
G72	66	9	year" laid down in the Education Act 1944. A similar	suggestion is made for courses for technicians. Since the
D07	98	8	imply that the proof texts were known by Abaye. Our	suggestion is substantiated by the fact that the comment on
H02	198	3	himself or for members of his family; (**=5) the	suggestion made in certain international organizations that
B02	162	5	has lost no time in taking up President Kennedy's	suggestion, made on Wednesday, that such a meeting should be
E28	148	3	have become entangled with other considerations. The	suggestion now made need have no party political implications
N07	60	2	from the mine. No place for boys to play." The	suggestion of a pout puckered the boy's face. "You talk like
K21	82	8	"ØGracias," acknowledged the white-armed one, a	suggestion of a smile returning to his lips. "ØAdios,

of modern spoken English.

This part of the London-Lund Corpus consists of 34 "texts", each of 5000 running words, thus totalling some 170,000 words of genuine face-to-face conversation produced by a number of British speakers in a variety of situations. The prosodic analysis includes such basic distinctions as tone unit, nucleus, booster, onset and stress. In the production of the printed version of the Corpus, which is intended for "manual" analysis, great care has been taken to ensure easy readability, as the following extract shows:

B 103 so ☆‖any time in JULÝ■ 104 ‖and☆ ÀUGUST■ 105 ‖but [ə:] + · +

A 106 ☆(- - a hiss-whistle)☆ +‖YĚS■+

>B 105 △not too 'far into 'August if ☆PÒSSIBLE■☆ - 107 ‖ŎTHERWISE■ 108 I'll be ‖stuck until about [ðĭ:]

A 109 ☆‖NŎ■☆

>B 108 △twentieth · [ə] I'm ‖HÔPING■ 110 to ‖get into SPĀIN■ · 111 from a‖bout the △twenty- · ▵ÈIGHTH of ÁUGUST■ 112 «to» un‖til about the △twentieth or

· ☆ ▵something ofthat kind of SEPTÈMBER■ ☆ 113 but

A 114 ☆‖YÈAH■☆

>B 113 ‖[ʌðəw] a‖part from ▵THĀT■ · 115 I'll be at ‖HÔME■ 116 and al'‖though I'll be doing CSĆ ▷stuff■ 117 and ‖that kind of THĪNG■ 118 ‖I can always 'put it on one ☆SĺDE■☆ 119 and ‖get on with the PÀPER■

A 120 ☆‖YÈAH■☆ 121 [ə:] you ‖see the △ŎTHER ▷man■ 122 ‖CHŎMLEY■ 123 ‖ought · ‖ought · ‖ought △ÀLSO■ 124 to have · ‖got his in on TĺME■ 125 and I SUS'‖PÈCTED■ 126 ‖ĂLWAYS■ 127 that De‖laney would be LÁTE■ · 128 that ‖Chomley would be on TĺME■ 129 and that ‖this would · produce a nice '△STÀGGERING■ 130 of · of their ar‖rival on your △DÈSK■ ☆-☆ 131 [ə:m] ‖now it looks as if they they both

B 132 ☆[m]‖[hm̀]■☆

>A 131 AR▵RÌVE■ 133 [ə] I ‖think that we ▵mustn't worry too ▵much A▵BŎUT THĺS■ 134 ‖we we ‖make it △perfectly clear that ▵papers must be in on the ▵first of ▵MĂY■ ☆-☆ 135 [ə:m]

B 136 ☆[m]‖[hm̆]■☆

>A 135 · [?ə ?ə:] ‖and [ə] I △don't want to [ə:] ‖you KNÓW■ 137 ‖run ourselves out of an external EXÁMINER■ 138 by ‖your SĀYING■ 139 [ə] oh to ‖hell with △THĪS■ 140 for a ‖GÁME■ 141 I'm ‖not going to have my summers · buggered up in ☆this kind of △WÀY■☆

The parallel version in machine-readable form provides easy access for the scholar who wants to make use of a computer.

THE BOOK

The printed version, which also includes presentation of the two "Survey" projects, a complete description of the corpus, a list of symbols, information about the speakers, and a description of the computer tape version, will shortly be published with the following title: *A Corpus of English Conversation*, edited by Jan Svartvik and Randolph Quirk (*Lund Studies in English*, Lund: CWK Gleerup). The book, which is a necessary tool for the user of the magnetic tape, can conveniently be ordered on the form at the bottom of this page. It cannot be obtained through ICAME but only from the publishers and their representatives.

THE TAPE

The computer tape is distributed by ICAME. For some technical information, see the order form at the end of the newsletter. Sound tapes of the material cannot be distributed.

SURVEY OF SPOKEN ENGLISH REPORT

The following report is available from the Survey of Spoken English, Porthuset, Allhelgona kyrkogata 14, S-223 62 Lund, Sweden: Cecilia Thavenius and Bengt Oreström, eds., *Konkordanser. Föredrag från 2:a svenska kollokviet i språklig databehandling i Lund 1979*. The report contains papers on concordancing (all of them in Swedish) by Rolf Gavare, Jonas Löfström, Benny Brodda, and Sven Norén.

ORDER FROM (BOOK)

I/We wish to order _____ copy/copies of Svartvik & Quirk, *A Corpus of English Conversation*, Lund Studies in English.

Name: _____

Address: _____

_____

(Return to: Gleerup/Liber Publishers, P.O.Box 1205, S-221 05 Lund, Sweden)

REPORT FROM A SYMPOSIUM ON GRAMMATICAL TAGGING OF ENGLISH TEXT CORPORA

An international symposium on "Grammatical Tagging of English Text Corpora in Machine-Readable Form" was held at Bergen on March 29-30, 1979. The symposium, which was financially sponsored by the Norwegian Research Council for Science and Humanities and the Universities of Oslo and Bergen, was arranged as part of the work within ICAME. It was attended by 37 participants from 10 countries.

The background to the symposium was the realization that corpora of (unanalyzed) natural-language texts are insufficient for many types of linguistic investigation, coupled with the discovery that linguists in different parts of the world had embarked on projects of grammatical tagging, seemingly unaware of each other's work and in some cases applying different systems of analysis to exactly the same material. During the Bergen symposium representatives from different projects had an opportunity to describe their work and profit from each other's experiences.

It is impossible to adequately summarize the papers and discussions. Wherever feasible, references will be made to publications giving detailed information on the particular projects.

Randolph Quirk (University College London) gave an introductory lecture on "The Place of Corpus Study in English Language Research". He emphasized the special features of the new corpora compared with the sources of material used by traditional grammarians such as Jespersen and Poutsma. In particular, the new corpora have been systematically compiled to represent a broad range of text types. They are further intended to be subjected to "total accountability" rather than to analysis of selected features. Quirk, who in his talk also touched on the relationship between corpus and elicitation, has recently dealt with these matters in a joint article with Jan Svartvik, "A Corpus of Modern English", in H. Bergenholtz and B. Schaeder, eds., *Empirische Textwissenschaft: Aufbau und Auswertung von*

Text-Corpora. Königstein/Ts.: Scriptor Verlag, 1979.
(This is the final title of the book which was announced in *ICAME NEWS* 1.)

If the preceding talk dealt with general linguistic matters, the particular uses of the computer in linguistics were taken up in brief contributions by Alvar Ellegård (University of Gothenburg) and Geoffrey Leech (University of Lancaster). Ellegård emphasized the importance of the computer in handling large bodies of data and relieving the linguist of much routine work, whereas Leech focused his remarks on the special advantages and possibilities offered by computer corpora and the need for cooperation in computer corpus work.

W. Nelson Francis (Brown University) presented the system which has been used in the recently completed tagged version of the Brown Corpus (cf. p.1 above). The system, which is essentially that outlined in B.G. Greene and G.M. Rubin, *Automatic Grammatical Tagging of English* (Department of Linguistics, Brown University, 1971), involves the assignment of one of 80 tags to each word in the material, through a combination of automatic procedures (dictionary look-up, suffix list look-up, context frame rules) and manual pre- and post-editing. The Brown Corpus tagging project is described in a paper by W. Nelson Francis, "A Tagged Corpus: Problems and Prospects" (forthcoming) and in the manual mentioned on p. 2 above.

Henry Kučera (Brown University) reported on results from studies of the tagged Brown Corpus in his talk on "The Frequency of Grammatical Classes in the Brown Corpus". Statistics were given for the frequency of individual tags (singular common noun, plural common noun, singular proper noun, etc.) as well as for major classes such as nouns, pronouns, verbs, etc. The latter were also ranked and compared with the frequencies in a Czech corpus. Word-class distribution across genres was further studied in a way which revealed the varying degree of "contextuality" of the major tag classes.

Alvar Ellegård (University of Gothenburg) described his analysis of portions of the Brown Corpus. This very detailed system, which, in contrast to that used at Brown University, does not involve any automatic procedures, has already been presented in this newsletter (*ICAME NEWS* 2, pp. 3-7).

Jan Aarts (University of Nijmegen) gave a report on "Grammatical Tagging in the Dutch Computer Corpus Pilot Project". The system is being implemented on a corpus of modern English texts assembled in Holland. It involves the manual assignment of a four-digit code to each word in the text and includes word-class labels comparable with those used in the Brown University project (the first two digits) as well as boundary markers (the last two digits). Categorial and functional constituents are derived from the four-digit code by a series of algorithms. The system has been adapted from J. van Bakel, *Automatische Syntactische Analyse van Nederlandse Teksten* (Computer Centre, Katholieke Universiteit, Nijmegen, 1970). Information on the Dutch project has been given in a paper by Jan Aarts on "Syntactic Coding of a Computer Corpus", presented at the 5th International Congress of Applied Linguistics, Montreal, August 20-26, 1978. The system of analysis is described in detail in a *Manual for Coders*, which is available on request from: Jan Aarts, Department of English, University of Nijmegen, Holland.

Rudolf Filipović (University of Zagreb), who was unfortunately prevented at the last moment from attending the symposium, submitted a paper on "The Grammatical Tagging of the 'Zagreb Version' of the Brown Corpus". In the Zagreb project about half of the Brown Corpus has been selected and translated into Serbo-Croatian, with the object of providing a source of data for contrastive analysis. The text is tagged manually according to a system in part reminiscent of Ellegård's and in part similar to that of the Dutch project. Words are assigned a four-digit code corresponding to part of speech (the first two digits), function of words or phrases in clauses (the third digit), and function of clauses in the sentence (the fourth digit), though the last two digits are only used with the first word of a syntactic constituent. Information on the Zagreb project has been given in pub-

lications by Filipović from the Serbo-Croatian-English Contrastive Project.

Jan Svartvik (University of Lund) gave an outline of the plans for the grammatical analysis of the London-Lund Corpus of spoken British English. The plans include semi-automatic word-class tagging similar to the Brown model as well as higher-level syntactic analysis. In his talk Svartvik touched on the particular problems of tagging spoken material, e.g. those posed by having the tone unit rather than the sentence as the basic element. The projected system is described in Jan Svartvik, "Tagging Spoken English" (forthcoming). See further the information on the London-Lund Corpus, pp. 6-8 above.

Mamata Nakra (Maisonneuve College) gave a talk on "Grammatical Tagging of Journalistic Prose" based on her work on newspaper material from the Brown Corpus, presented in her thesis on the topic. Nakra's system has not yet been implemented computationally.

Viljo Kohonen (University of Turku) described the CHITAB program, which he has developed in cooperation with Jussi Salmela. The program operates on a coded version of the text (manually assigned) without direct access to coding and text at the same time. It has been used in Kohonen's recently completed thesis, *On the Development of English Word Order in Religious Prose around 1000 and 1200 A.D.* Publications of the Research Institute of the Åbo Akademi Foundation, No. 38. Åbo 1978, and is described on pp. 223-227 of his work.

Claus Faerch (University of Copenhagen) reported on the grammatical analysis of a corpus of learners' language collected in Denmark and consisting of English as spoken and written by Danes. The tagging is restricted to the assignment of word-class labels by semi-automatic techniques along the Brown model. Faerch touched on the particular problems caused by the learner-language material. Is it possible to characterize learner-language as a

system? If not, can you write rules for the assignment of tags? The Copenhagen project is described in Faerch's report on "Computational Analysis of the PIF Corpus of Learner Language", *PIF Working Papers*, No. 1, Department of English, University of Copenhagen, 1979.

Dirk Geens (University of Leuven) was the only one of the participants who gave a report on automatic syntactic analysis, based on his recently completed doctoral thesis on the topic. Geens' analysis, which has been implemented on the Leuven Drama Corpus (described in *ICAME NEWS* 2, pp. 7-9), included a "syntactic recognition procedure that was mainly used to produce a rough characterization of the apparent syntactic structure of each sentence" and a "full syntactic analysis procedure", both of which are too complex to be summarized here. Geens also dealt with some problems of semantic analysis.

In a lucid and relevant introduction to the final discussion period, Sture Allén (University of Gothenburg) presented a taxonomy of tagging systems based on the type of material (running text, sorted text, linked network), purpose of analysis (lexical, grammatical, communicative), and tagging technique (off-line encoding, on-line encoding, interactive procedure, automatic procedure). Through Allén's paper the work at the symposium was placed within a more general framework of computational-linguistic analysis.

The contributed papers presented a variety of tagging projects, from the point of view of the material (cf. Francis, Svartvik, Faerch) as well as aim (cf. Francis, Filipović, Geens) and technique (cf. Francis, Ellegård, Geens). An illustration of different ways of tackling the same material was given by four treatments of an extract from the Brown Corpus: the Brown model, Ellegård's and Filipović's systems, and the Dutch system (Jan Aarts was kind enough to provide comparative material, though

the text was not part of the corpus analyzed in the Dutch project). Unfortunately, space does not permit the reproduction of this material here.

The variation in the ways of handling the same material raises the question why a particular system is chosen. It was clear from the discussion that different projects had different aims, as already hinted at in the preceding paragraph, which partially explains the varying approaches (Naturally, these are also due to such more trivial matters as available funds, personnel, equipment, etc.) The Brown analysts were concerned with developing a source of data as neutral as possible with respect to future applications and were therefore content with a fairly uncontroversial, traditional, "surface" analysis, whereas Ellegård had the more specific aim of revealing the syntactic features of four categories of English texts and therefore considered it necessary to perform a fuller analysis. The varying aims in other cases (Faerch, Filipović, Geens) should be immediately recognizable to the reader.

The linguistic differences between the systems should, however, not be exaggerated. Most of them use traditional categories (parts of speech, parts of the sentence, clause types), though the delicacy of analysis and the labels may differ. The question of standardization of categories and labels was raised in the discussion. It was agreed by almost everybody that standardization is impossible, or even undesirable, in view of the varying aims, ambitions, and resources of projects. Instead, it was pointed out that it is desirable to have systems which are convertible to some extent and always to provide explicit descriptions of the systems used.

In conclusion (to take up just one further matter from the discussion), it was agreed that the exchange of information between researchers must be improved. To partially solve this problem, it was proposed that a follow-up symposium should be held in two years, if possible.

THE OXFORD ARCHIVE OF ENGLISH LITERATURE

The Oxford Archive of English Literature has the following aims:

(a) to establish and maintain a central repository for English literature in machine-readable form;

(b) to make available texts from it to researchers in the field of literary and linguistic computing, free of charge (other than for transport etc.).

In contrast to ICAME, the Oxford Archive focuses on printed literary texts, though other material is also available, as appears from the following list.

1. Complete English texts

Anonymous	Arden of Faversham; Contention of York & Lancaster; Famous Victories of Henry V; Taming of a Shrew; Troublesome Reign of King John, True Tragedy of Richard III; Thomas of Woodstock
	Sweet Gooseberries
Akenside, Mark	Pleasures of the Imagination
Barnes, Barnabe	Sonnets
Beckett, Samuel	Ping; Bing; Lessness; Waiting for Godot
Berryman, John	Dream Songs
Bullokar, William	Three pamphlets on grammar
Chesterfield, Earl of	Case of the Hanover Forces ...
Coleridge, Samuel Taylor	Complete poetical works (ed. E.H. Coleridge)
Cooper, Giles	Everything in the garden
Cowper, William	The task [in preparation]
Crane, Ralph	Rejoinder to a bill of complaint
Dickens, Charles	A Christmas Carol
Dylan, Bob	Published songs 1962-69
Eliot, Thomas Stearns	Complete poems and plays; Poems 1909-35
Erasmus (trans. Hervet)	De Immensa Dei Misericordia
Fielding, Henry	Joseph Andrews; Miscellanies; Shamela
Fleming, Ian	Goldfinger
Fletcher, John & Massinger, William	Sir John Van Olden Barnavelt
Fletcher, John	Demetrius and Erianthe
Frost, Robert	Selected verse
Gaskell, Elizabeth	Selected contributions to Frasers 1851-65
Gower, John	Confessio amantis
Graves, Robert	Complete poems
Greene, Robert	Friar Bacon and Friar Bungay
Hervey, Thomas	Letter to Sir Thomas Hanmer, Bart
Hopkins, Gerard Manley	Complete verse
Johnson, Samuel	London; The vanity of human wishes
Jonson, Ben	Pleasure reconcild to vertue (a masque)
Keats, John	Poetical works
Kydd, Thomas	Spanish Tragedie; Cornelia
Lawrence, David Herbert	St Mawr

Mansfield, Katherine	Selected short stories
Manwaring	A seaman's glossary
Middleton, Thomas	A game at chess (3mss); The Witch; Song in several parts
Munday, Thomas et al	Sir Thomas More (parts)
O'Casey, Sean	Plough & the stars; Juno and the paycock; *Shadow of a gunman*
Randolph, Thomas	Aristippus; Conceited Pedler; Praeludium; Drinking Academy; Fary Knight
Shakespeare, William	Complete works (First folio and assorted quartos)
Spenser, Edmund	Minor Poems; Faerie Queene
Tourneur, Cyril	Revengers Tragedy
Woolf, Virginia	A haunted house and other stories
Wordsworth, William	Lyrical Ballads (1798)
Wyatt, Thomas	Poetical works

2. Reference Texts

Advanced Learner's Dictionary	[headwords; new complete edition in preparation]
Shorter Oxford Dictionary	[headwords & syntactic codings only]
Thorndike-Lorge word count	
General Catalogue of the OUP	[in preparation]
African Encyclopaedia	[in preparation]
Oxford Dictionary of Quotations	[in preparation]

3. *Corpora*

Brown Corpus
Gill Corpus 1 (to be described in a
Gill Corpus 2 later issue of *ICAME NEWS*)
Lexis (spoken usage)

4. Samples

Modern prose	(10 approx. 2000 word samples from novels by Huxley, Snow, Lewis, Waugh, Greene, Wells, Murdoch, Wyndham, Maugham, and Woolf)
Civil War polemic	(19 3000 word samples from pamphlets by Milton; 15 3000 word samples from pamphlets by Hall. 'SMECTYMNUUS' et al.)
Dedications	transcribed by Ralph Crane

A more complete list of material available, as well as detailed information on price and conditions of use, can be obtained from: Louis Bernard, The Archive, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, England.

THE LLCC ARCHIVE

The Literary and Linguistic Computing Centre at the University of Cambridge (LLCC) has available a large number of texts in machine-readable form, representing a variety of languages. Like the Oxford Archive, the focus is on printed literary texts. The following is an incomplete list of the English texts currently available.

Bayes, H.	Transcribed Speech of a Small Child (unpubl.)
Dudley, Third Lord North	Collected Poems
Dudley, Fourth Lord North	Collected Poems
Eliot, George	Silas Marner; Middlemarch; Daniel Deronda (other novels to follow)
Eliot, T.S.	The Complete Poems and Plays (London, 1969)
Graves, Robert	Complete Poems, first published editions
Graves, Robert	Claudius; Claudius the God (other prose works to follow)
Hopkins, Gerard Manley	The Wreck of the Deutschland
Mansfield, Katherine	Collected Stories (selected stories only) (Constable, 1968)
Spenser, Edmund	Færie Queene, 2 vols., edited by J.C. Smith (Oxford, 1902)
Spenser, Edmund	Minor Poems
	A Systematic Sign Language, a dictionary of deaf and dumb language signs
Wyatt, Sir Thomas	Collected Poems
Woolf, Virginia	A Haunted House, and Other Stories, edited by L. Woolf (London, 1967)

A more complete list as well as detailed conditions of use can be obtained from: John Dawson, Literary and Linguistic Computing Centre, Sidgwick Site, Cambridge CB3 9DA, England.

EDITORIAL NOTE

Further ICAME newsletters will appear irregularly and will, for the time being, be distributed free of charge. The Editor is grateful for any information or documentation which is relevant to the field of concern of ICAME.

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 and return the form below (unless you did so after receiving *ICAME NEWS* 2).

(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

To obtain material you must enclose a cheque (bank draft, cashier's cheque) with your order made out in Norwegian currency (or the equivalent in English pounds or U.S. dollars) to: The Norwegian Computing Centre for the Humanities, Bergen, Norway.

(Use of the material for research by commercial institutions may be possible on the same conditions. Commercial institutions are, however, asked to contact us before ordering any material.)

(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
London-Lund:Text	1 1200 ft.	150	1 1200 ft.	200	1 1200 ft.	200

☐ 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.

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 _____

TAGGED VERSION OF THE ENGLISH CORPUS TAPES

Order Form

Mail to: Henry Kucera
TEXT RESEARCH
196 Bowen Street
Providence, R.I. 02906, USA

Date:

1. Please send ___ copy(ies) of the Tagged Corpus of Present-Day American English (each copy contained on two magnetic tapes) on 9-channel tapes in the following density (check one):

__ 800 BPI __1600 BPI __6250 BPI

Code: ____ASCII ____EBCDIC
(If blank, EBCDIC will be supplied)

2. Instead of the above, send 7-channel tapes in ____ BPI density, in BCD code. I understand that 7-channel tapes, because of processing complications, incur an additional charge of $25.- per each complete set of the Tagged Corpus.

3. Send tapes via __Surface transport prepaid __Air freight collect to the following address (please give street address and Zip code for UPS delivery):

4. Send ___extra copies of the Manual at $6.50 each (one copy supplied free with tapes)

5. Payment:

__ Please send invoice
__ I enclose a check, payable to TEXT RESEARCH, in the amount of
__ $250.- (for non-profit organizations)
__ $350.- (for others)

If you wish the rate for non-profit organizations, please give the name and address of the organization :

6. I agree to the following conditions:

I understand that the Tagged Version of the English Corpus is protected by copyright and subject to the provisions of the U.S. copyright law. No copy of the tapes will be made without the written permission of one of the copyright holders, W.N. Francis or Henry Kucera.

Signature:
Print or type name:
Position:

N.B. The Tagged Brown Corpus cannot be obtained from Bergen!