

## The coming of ICAME

*Geoffrey Leech, University of Lancaster*

*Stig Johansson, University of Oslo*

### **Abstract**

*Two of the founders give a brief account of the origin and early development of the International Computer Archive of Modern English (ICAME), focusing on the prehistory of ICAME, the founders' meeting in Oslo in February 1977, and the period up to the first ICAME conference in Bergen in March 1979.*

### **The prehistory of ICAME (by Geoffrey Leech)**

In a general sense, the prehistory of ICAME must include the foundations of English corpus linguistics as laid by Randolph Quirk, who established his Survey of English Usage at University College London in 1959–60, and by W. Nelson Francis, whose brainchild was the Brown Corpus,<sup>1</sup> compiled at Brown University in 1961–64. Although English language corpora of a limited kind had been used before, Quirk and Francis both recognized the value of what is now often called a ‘reference corpus’: a corpus as a general linguistic research resource, representative of a wide, if not comprehensive, range of genre or text types, to be used by ‘all comers’. The key difference between Quirk’s and Francis’s inspiration was that, for Quirk, interest in the spoken language was uppermost, and, in those days of primitive computers, the computational handling of speech was scarcely to be dreamed of. So the Survey of English Usage corpus remained in huge metal fireproof cabinets in two or three rooms in Foster Court – in University College London – and the world of scholarship had to make its pilgrimage there to consult its contents. For Francis, on the other hand, the inspiration lay in the power of computers: if a corpus could be created on computer, it could be copied and distributed so that the world of scholarship did not need to make its pilgrimage: the corpus would come to the scholar.

The early days of the Brown Corpus were told by Nelson Francis in an amusing after-dinner speech at the fifth ICAME conference at Windermere, England, on 21 May 1984.<sup>2</sup> The flavour of it can be captured from these opening paragraphs:

You probably can't see it from where you sit, but some of you may have noticed that I am wearing a tie clip in the shape of a monkey wrench – or what I believe is called an 'adjustable spanner' in the curious dialect of this country. The story behind this peculiar piece of jewelry goes back to the early 60s, when I was assembling the notorious Brown Corpus and others were using computers to make concordances of William Butler Yeats and other poets. One of my colleagues, a specialist in modern Irish literature, was heard to remark that anyone who would use a computer on good literature was nothing but a plumber. Some of my students responded by forming a linguistic plumber's union, the symbol of which was, of course, a monkey wrench. The husband of one of them, being a jewelry manufacturer, had a few of these clips made. I cannot say that they have become collectors' items, but I would certainly not part with mine.

I later encountered that colleague on some social occasion, and he had the grace to say "Ah, Nelson, me bhoy, it was not you I was after callin' a plumber; it was them other fellows like Henry Kučera." – I should point out that much reading of Sean O'Casey had had a strange effect on his speech. I don't think he is genuine Irish; if so, he's the only Irishman I ever met named Kraus. People are more familiar with computers nowadays, and perhaps not so hostile as my colleague David O'Kraus.

This was not the only kind of hostility encountered by corpus linguists in those days, and Nelson elsewhere recounted his brush with Chomsky's leading disciple, who told him that collecting a corpus was a complete waste of time.<sup>3</sup>

To be absolutely precise, the place and date of the founding of ICAME can be fixed as at Oslo, on 12<sup>th</sup> February 1977. In a sense, however, the story of how ICAME came into being can be traced back to 1969, the year when, as a young academic, I moved from University College London (UCL), where I had been a very junior colleague of Randolph Quirk, to the brand-new University of Lancaster (founded in 1964). If I may beg readers' patient indulgence, then, I will take you back to a room on that new campus in the autumn of 1969, where a small group of young and inexperienced academics sat round a table to discuss how we were going to put Lancaster on the map for research on the English language. We were the 'modern linguists' of a largish Department of English, and I had just been appointed to a readership to lead this section of the Department.

Coming, as I did, from the department of Quirk, whose work on the Survey of English Usage had impressed me at UCL, it was perhaps not surprising that

my first suggestion was “Let’s try to build a computer corpus of English”. I had met Nelson Francis in London, where he had famously presented RQ with a copy of the Brown Corpus, in the form of a weighty magnetic tape, with the words “Habeas corpus!”. It seemed an obvious yet novel plan to compile a British counterpart of this pioneering American corpus, so I opened a transatlantic correspondence with Nelson, who welcomed the idea with enthusiasm, and readily gave help and encouragement. I am particularly grateful, in retrospect (for the benefits were not fully evident then), that he advised me to imitate the composition of the Brown Corpus in all possible detail, so that the idea of *comparable corpora*, which came to fruition in the Brown family of corpora, as well as in ICE and ICLE, can be said to have started in a letter that Nelson wrote to me on 26 November, 1969:

I think the Lancaster Corpus should be as isomorphic as possible with the Brown Corpus – not because we had a monopoly of wisdom in making selections (though RQ was one of our consultants!) but because the possibility of comparative studies, especially statistical ones, is worth making the effort for close correspondence.

After some fruitless funding bids, we were eventually awarded the princely sum of £3000 by the publishers Longman, and the Lancaster Corpus, which eventually would become the Lancaster-Oslo/Bergen Corpus, was soon under way.<sup>4</sup> Already the plan had become more grandiose: the intention was to set up an electronic archive of corpora known as CAMET (the Computer Archive of Modern English Texts), consisting initially of the Brown Corpus, the embryonic Lancaster Corpus, and the Survey of English Usage corpus in London, which already contained much spoken material.

However, these grand plans ran into the ground. Although further small amounts of money were obtained, costs escalated, and three major difficulties confronted us: computational inexperience, primitive computing facilities, and copyright problems – of which the last proved most intractable. Following the Brown experience, we wrote to the copyright holders of the 500 text samples of which the corpus would consist, explaining the academic excellence of the project, pleading poverty, and begging the copyright holders to grant us permission without payment of a fee. However, British publishers and agents proved on the whole more hard-hearted than the American ones, and attempts to lobby national bodies, such as the British Council, the Philological Society, and the Publishers’ Association met with sympathy but little success. Finally, by 1976, I despaired of the whole project.

At this point the CAMET project was rescued by Stig Johansson, who, as a protégé of Jan Svartvik, had won a Leverhulme Fellowship to come from Lund to spend a year (1973–74) at Lancaster. Stig's wife Faith-Ann, who was without a job at Lancaster during that year, found herself employed far below her scholarly capabilities, in the most depressing donkey work on the corpus, typing and checking data on paper tape, as a means of inputting texts to the computer. Where errors had occurred, which was often, these had to be painstakingly corrected by splicing the tape. But somehow Faith-Ann found some interest in the work, as did also her husband. When Stig returned to Scandinavia and obtained a docent's post at Oslo, he wrote to me offering to try to finish off the corpus. It can be imagined what a relief that was. We transported all the corpus materials to Oslo, and eventually CAMET became ICAME, and the Lancaster Corpus became the LOB Corpus.

How did this transformation take place? Part of the thinking behind ICAME was that we needed to impress the British copyright holders, mostly publishers and agents based in London, who had not been sufficiently impressed by the entreaties of an obscure new provincial university in the north of England. If they were to receive a request from the coordinating secretary of a high-sounding international organization based outside the UK, they might see the inclusion of their text in an archive as more of an honour to them, than a favour to the applicant. This stratagem was eventually successful: after the founding of the International Computer Archive of Modern English, the completion of the corpus in Norway was at length achieved in 1978, through the commitment of the team led by Stig, which included Jostein Hauge and Knut Hofland at the Norwegian Computing Centre for the Humanities at Bergen.<sup>5</sup>

The actual founding of ICAME took place at Oslo after a flurry of urgent letters had passed between Nelson Francis, Stig Johansson, Jan Svartvik and myself. (E-mail, had it been available, would have helped immensely!<sup>6</sup>) We all recognized the value and timeliness of establishing some kind of international organization for archiving, documenting and distributing computer corpus resources. This needed to be done in a hurry, as funding was needed to complete the project in Norway, and I was due to leave Europe for a three-month visit to China in February 1977. I was going to lead the first academic delegation<sup>7</sup> from the UK to the People's Republic of China, an unusual destination at that time, when regular communication with colleagues in the West was likely to be difficult. Nelson and Jan were full of enthusiasm for the project. The Brown University project (with Henry Kučera) for the POS-tagging<sup>8</sup> of the Brown Corpus was well advanced, and Jan was nearing the completion of his computerized version of spoken texts from the Survey of English Usage – the London-Lund

Corpus, as it became. This corpus, together with Brown and LOB, was to form the nucleus of the “international archive”.

The University of Oslo was the most convenient place for our meeting, as happily, Nelson was at the time a visiting professor at another Norwegian university, the University of Trondheim. Jan’s journey from Sweden and mine from the UK were obviously manageable. A fifth member – Arthur Sandved – gave us his backing as the Head of Stig’s department, and completed the founding group.

The flurry of letters leading up to the Oslo meeting of 10–12 February dealt with a range of issues. For example, there were a number of suggestions for extending the circle of invitees and contacts, but it was mainly the imminence of my departure for China that persuaded us to act quickly with a small group, rather than to involve a larger group of interested scholars at this stage.

In a more light-hearted vein, there was discussion of the suitability of the name ‘ICAME’. Jan ended his letter to me of 19<sup>th</sup> January with the handwritten note:

Isn’t there a better name? ARCHENG/ ARCHENGLISH/ ENGARCH/ COMPARCH/  
COMPENG/ ENGCOMP/ INTERENG/ .... (obviously not!)

Nelson ended his first letter to Stig (on 26<sup>th</sup> January) as follows:

Meanwhile let me say I am delighted that you are willing to help Geoff out and to serve as co-ordinator and general secretary of the projected ICAME (which will inevitably have to be succeeded by ISAW, though perhaps ICONQUERED had better be left for a future generation).

In so far as one can separate the roles of the four correspondents, I seem to have suggested the formation of the international archive, Jan proposed the founding meeting, Nelson suggested that there be a meeting in Oslo, and Stig, as co-ordinator, arranged and hosted the whole event.

At the end of the meeting, we all signed a document announcing ICAME to the world and inviting interested parties to get involved. This document is reproduced in Figure 1.

## UNIVERSITY OF OSLO

ENGLISH DEPARTMENT  
TELEPHONE 466800



P.O. BOX 1003  
BLINDERN, OSLO 3, Norway

The undersigned, meeting in Oslo in February 1977, have informally established the nucleus of an International Computer Archive of Modern English (ICAME). The primary purposes of the organization will be:

- (1) collecting and distributing information on English language material available for computer processing;
- (2) collecting and distributing information on linguistic research completed or in progress on the material;
- (3) compiling an archive of corpuses to be located at the University of Bergen, from where copies of the material could be obtained at cost.

Initially, the archive will include the Brown University Corpus of American English, the complementary corpus of British English texts which has been under preparation for a number of years at the University of Lancaster, England, and the spoken texts of the Survey of English Usage (University College London), which is now being transferred to magnetic tape for computer processing at the University of Lund, Sweden. As needs and opportunities arise, the archive will be extended to include other corpuses.

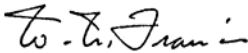
One of our main aims in establishing the organization is to make possible and encourage the coordination of research effort and avoid duplication of research.

We hope that, as someone interested in this type of research, you will join us in this enterprise. Specifically, we would like to receive information on the following points:

- (a) Are you currently engaged in computerized linguistic research?
- (b) Do you know of any scholars (other than those on the attached list) who would be likely to be interested in the organization?
- (c) Have you compiled a corpus which you would be willing to share with others? If so, under what conditions?
- (d) Do you know of other researchers who have compiled corpuses of Modern English which might be included in the archive?

Please, address your responses to our coordinating secretary, Dr. Stig Johansson, University of Oslo.

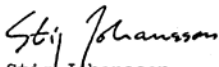
Oslo February 12, 1977



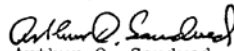
W. Nelson Francis  
(Emeritus Professor of  
Linguistics, Brown University)



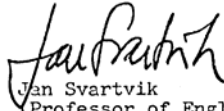
Geoffrey Leech  
(Professor of Linguistics and  
Modern English Language,  
University of Lancaster, U.K.)



Stig Johansson  
(Docent in English Language,  
University of Oslo)



Arthur O. Sandved  
(Professor of English Language,  
University of Oslo)



Jan Svartvik  
(Professor of English,  
University of Lund,  
Sweden)

Figure 1: The document announcing the founding of ICAME

## *The early history of ICAME (by Stig Johansson)*

### *Completing the LOB Corpus*

As Geoff has explained, the beginning of ICAME was intimately connected with the work on the LOB Corpus. The immediate task for me was to raise funds for completing the corpus. I was glad to get a grant from the Norwegian Research Council for Science and the Humanities and also some funding from my own department. Most significantly, I established cooperation with the Norwegian Computing Centre for the Humanities and its Director Jostein Hauge. I received a scholarship from the Centre for a four-week visit in May–June 1977 to learn more about computing in the humanities. One of the teachers was a young computing engineer, Knut Hofland, who became my life-long friend and collaborator. With his computational expertise and his willingness to learn about the way linguists think (and his exceptional personal qualities), his work was crucial not only for the completion of the LOB Corpus but also for the development of ICAME.

As getting permission from copyright holders had turned out to be so intractable, this was the first problem to be addressed. I got in touch with the top expertise on copyright in Norway and was advised that the only way to proceed was to secure the written permission of the copyright holders. This was an arduous task, because a great many different authors and publishers were involved. As quite a long time had passed since the publication of the 500 texts (1961), it was often difficult to trace the copyright holders. After this job had been done, each of these had to be contacted. Several hundred letters were sent out asking for permission to use the texts for research. According to my files, 160 letters were sent in June 1977 alone.

Fortunately, most of the answers were positive. One author was so inspired that he wrote a collection of poems, “Songs of Norway”, which he enclosed with his response. But there were also some negative replies, and in a few cases we were asked to pay for the use of the texts. One copyright holder asked for 40 pounds for a single text, and in a letter to Geoff in January 1978 I wrote: “Would you believe it that the Daily Telegraph wrote recently that they had let me off too easily and wanted 100 pounds?” Where there were problems of this kind, I wrote a new letter explaining more about the project and appealing to copyright holders to re-consider their decision. Follow-up letters also had to be written in the great many cases where we had not received a response. As a result of this work, we managed to keep the great majority of the texts brought from Lancaster, and only a small number had to be replaced. The new texts were converted to machine-readable form at the Bergen Centre.



The next task was to proofread the texts. I had a number of student helpers, but had to do most of this time-consuming work myself. Knut provided me with printouts of the texts, and I returned lists of errors, which were then corrected at Bergen. In view of the technical difficulties in entering the texts at Lancaster, there were a great many errors in some of the texts, and a lot of letters were exchanged between Oslo and Bergen. As there had been a rapid technological development, working conditions in Bergen were much better. Now there was no need to work with paper tapes, and the text could be seen and errors corrected on a computer screen.

Apart from the correction of errors, a good deal of work involved checking the consistency of coding of abbreviations, foreignisms, sentence boundaries, etc. This required a good deal of thought. Again Knut was helpful in sending lists for consistency checking, and the staff at the Centre entered the necessary corrections. Finally, the manual (Johansson *et al.* 1978) for the LOB Corpus was written, building on manuscripts from Lancaster. The manual and the whole corpus were completed by the end of 1978. The rescuing of the LOB Corpus turned out to be successful and formed an auspicious start for ICAME and the Oslo-Bergen cooperation.

### ***Distribution of texts***

One of the tasks of the new organization was “compiling an archive of corpuses to be located at Bergen, from where copies of the material could be obtained at cost”. Here the Bergen Centre has been invaluable. Knut produced a revised version of the Brown Corpus with upper- and lower-case letters and other features which reduced the need for special codes and made the text more easily readable. The first product which was ready for distribution was a microfiche concordance of the Brown Corpus, first announced in December 1977. The availability of the Brown texts was announced in *ICAME News* 1 (March 1978), and the LOB texts were ready for distribution in 1979, as announced in *ICAME News* 2 (March 1979). Computer tapes for the London-Lund Corpus became available later in 1979, as announced in *ICAME News* 3 (October 1979).

The interest in ICAME and the material distributed through ICAME grew rapidly. Table 1 gives a survey of the distribution by country (institutions or individual researchers) up to 31 December 1985. A similar list was published in *ICAME News* 8 (May 1984), p. 2.

*Table 1:* The distribution of ICAME material by country  
(up to 31 December 1985)

Australia	4	Japan	10
Austria	1	Netherlands	6
Belgium	5	New Zealand	1
Canada	6	Norway	6
Denmark	6	Sweden	10
Finland	10	Switzerland	1
France	2	UK	27
Hong Kong	1	USA	18
Israel	2	West Germany	12
Italy	4		

In the beginning there were not many electronic corpora, and I was often confronted with problems in answering requests from private companies who wanted to get copies of the corpora for research and development projects. Because of the conditions under which copyright permission had been obtained, the general reply had to be that ICAME material was only available for non-profit research and was normally only distributed to academic institutions (or individual researchers connected with such institutions), but there were some marginal cases. After machine-readable corpora became more numerous, these requests gradually stopped.

Over the years the number of corpora distributed through ICAME, the form of the texts (raw text, tagged text, concordance), and the distribution format have changed greatly. To give an idea of the format of the material distributed, I summarize figures for the period 1991–95; see Table 2. Unfortunately, I do not have reliable figures for the first few years, but we know that the distribution format was initially limited to microfiche and magnetic tape.

*Table 2:* The distribution of ICAME material in the period 1991–95

Format	1991	1992	1993	1994	1995	1991 – 1995
Diskette units	79	61	55	91	23	309
CD-ROM	0	61	59	70	43	233
Cartridge/Video8	17	5	12	5	3	42
9-track mag tape	10	4	3	0	0	17
Microfiche	2	0	0	0	0	2

As we see from the table, there has been a complete change, and further changes were to follow.<sup>9</sup>

### ***The ICAME newsletter***

To simplify the running of ICAME, we started publishing the newsletter *ICAME News*. This was a modest publication giving basic information on the ICAME corpora and on how they could be obtained, but there was also some more general information on related projects, relevant conferences, etc. In the beginning I was responsible for writing the whole text, but gradually the publication grew and came to include articles, reviews, conference reports, etc. The first ICAME Bibliography was published in *ICAME News* 10 (May 1986).<sup>10</sup> By 1987 we felt ready to change the name to the *ICAME Journal*, to reflect the change of the publication, both in form and content. This was also when the first Advisory Board was announced, originally established at the ICAME Conference in 1986. This was the first step in the direction of greater formalization of ICAME.

For many years the newsletter/journal was distributed without cost to the subscribers. The number of subscribers grew rapidly and by 1989 there were close to 600.<sup>11</sup> Although the great majority were from Europe or North America, there were also subscribers from Australia, China, India, Israel, Japan, Korea, New Zealand, Qatar, South Africa, Taiwan, and Thailand. To keep the ICAME operation going, it became necessary to ask for a subscription fee, and then the number of subscribers declined drastically. All the issues of *ICAME News* and the *ICAME Journal* are now available on line; see the ICAME website (<http://icame.uib.no/>).

### ***The first ICAME conference***

As one of the main aims in establishing ICAME was “to make possible and encourage the coordination of research effort and avoid the duplication of research”, it was natural to arrange conferences where scholars with similar research interests could meet. What we now count as the first ICAME conference was originally a symposium arranged in preparation for the grammatical tagging of the LOB Corpus. The symposium was arranged at Bergen on March 29–30, 1979, and the theme was ‘Grammatical Tagging of English Text Corpora in Machine-readable Form’. There were 37 participants from 10 countries. See the report in *ICAME News* 3, pp. 9–14. Note also the photo on the ICAME website showing some of the key figures in the development of ICAME. Three of the talks (by Randolph Quirk, W. Nelson Francis, and Geoffrey Leech) are also available as audio recordings on the website. Although a full report is given in the newsletter, I will mention some of the contributions here, giving the full names of the contributors and their academic affiliation at the time of the conference.

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 had been systematically compiled to represent a broad range of text types. They were also intended to be subjected to “total accountability”, rather than to analysis of selected features. Quirk further touched on the relationship between corpus and elicitation, both of which have an important role to play in the study of language.

W. Nelson Francis (Brown University) dealt with the system used in the tagged Brown Corpus, which had just been completed at this time, and Henry Kučera (Brown University) reported from studies of the tagged corpus in his talk on ‘The frequency of grammatical classes in the Brown Corpus’. Jan Aarts (University of Nijmegen) gave a report on ‘Grammatical tagging in the Dutch Computer Corpus Pilot Project’. Jan Svartvik (Lund University) outlined the plans for the grammatical analysis of the London-Lund Corpus. He touched on the special problems of analyzing spoken material. Dirk Geens (University of Leuven) was the only one of the participants who gave a report on automatic syntactic analysis, as implemented on the Leuven Drama Corpus. He also dealt with some problems of semantic analysis.

Among other contributions was an informal talk by Geoffrey Leech (University of Lancaster), where he stressed the special advantages and possibilities of computer corpora and the need for cooperation in computer corpus work. It is especially significant that he looked ahead to what might happen in the future. A paper by Viljo Kohonen (University of Turku) presented a program devised in connection with his work on English historical texts. Claus Færch (University of Copenhagen) gave a report on the grammatical analysis of a corpus of learners’ language consisting of English as spoken and written by Danes. He 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? Studies of historical texts and learner language were only later to become major areas of research within the ICAME community.

In a lucid and relevant introduction to the final discussion period, Sture Allén (Göteborg University) presented a taxonomy of tagging systems based on the type of material (running text, sorted text, linked network), the purpose of analysis (lexical, grammatical, communicative), and the tagging technique (off-line encoding, on-line encoding, interactive procedure, automatic procedure). Through this 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 type of material as well as the aim and the tagging technique. The question of standardization of categories and labels was raised in the discussion. It was agreed by almost everybody that standardization was impossible, or even undesirable, in view of the varying aims, ambitions, and resources of the 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, it was agreed that the exchange of information between researchers must be improved. To further cooperation and the exchange of information, it was proposed that a follow-up symposium should be held in two years, if possible.

No one knew at this time that there would be a long series of ICAME conferences; see the list on the ICAME website. The first two meetings following the tagging symposium were small, with a limited number of participants (Bergen 1981 and Stockholm 1982), but they served to keep up the momentum. By the time of the Nijmegen conference in 1983, the number of participants was much greater, and to accommodate the conference report (guest-edited by Jan Aarts) we had to split *ICAME News* 8 (May 1984) into two parts. But it was not until the 1985 conference, arranged at Röstånga, Sweden, by Jan Svartvik and his team, that we started referring to a series of ICAME conferences; see the report from the Sixth ICAME Conference in *ICAME News* 10 (May 1986). We are now approaching the 30<sup>th</sup> conference, to be arranged in May 2009 at Lancaster. Further conferences are already at the planning stage.

### ***The role of the Bergen Centre***

The role of the Norwegian Computing Centre for the Humanities at Bergen cannot be stressed enough. From the beginning the Centre has been responsible for the management and distribution of texts (later also programs). It hosted the first two ICAME conferences (1979, 1981) and another one in 1989. It maintains the ICAME website. It has been responsible for the publishing of the newsletter/journal; for a long period Anne Lindebjerg has done an important job in preparing the computer files for printing.<sup>12</sup> The key figure over the years has been Knut Hofland, whose importance has already been referred to. In addition to what has been mentioned, he is the Technical Secretary of ICAME. He started and still maintains the Corpora list, currently with about three thousand subscribers across the world. Other key people at the Centre have been Jostein Hauge,<sup>13</sup> who generously supported ICAME from the beginning and until he left the post as Director of the Centre, and members of the secretarial staff, including Edle Burgess, Torill Revheim, and Kari Sørstrømmen.

### ***Moving ahead***

A lot has happened since the founders' meeting in February 1977. At that time we could not foresee the vast technological changes that were to follow nor the increasing interest among linguists in using corpora and studying language in use. The number, size, and range of corpora have increased immensely, and so have the sophistication and range of analysis tools and the number and diversity of uses and users. At the outset those using computer corpora were few and were considered to be outside the mainstream of linguistics; what is known as corpus linguistics has now been widely accepted. As one of the founding fathers put it in 1996, "corpora are becoming mainstream".<sup>14</sup> There is no sign that the development has slowed down since then. ICAME has no doubt played an important role in this development.

ICAME has changed over the years. After I gave up as coordinating secretary in 1996, the organization was formalized, with a Chairman and (later) a Constitution. At this time it also got a new name, the International Computer Archive of Modern and Medieval English (keeping the old acronym), to recognise the flourishing of historical corpus work, which was in large measure due to the inspired leadership of Matti Rissanen, who became the new Chairman.

Another way in which ICAME has changed is the increasing number of female members. The founders' meeting was all-male, no doubt reflecting realities in the academic world at the time. Now female members are abundant and influential. Among the earliest female ICAMERs are Antoinette Renouf (University of Birmingham, now at Birmingham City University) and Nelleke Oostdijk (Nijmegen), both of whom have been part of the ICAME community since the early 1980s. Antoinette is the present Chair, and Nelleke is a member of the Board. Other key female members in the early period include Karin Aijmer (Göteborg), Gunnel Tottie (Zürich, previously at Uppsala and Lund), and Merja Kytö (Helsinki, now at Uppsala), who is Secretary of ICAME and one of the editors of the *ICAME Journal*. The other editor of the journal is also female:<sup>15</sup> Anna-Brita Stenström (still active, though retired from her post at the University of Bergen). It is a sign of the times that two thirds of the present Board are female.

In this paper we have focused on the early years and have only referred in passing to later developments. Those who want to learn more are recommended to consult the ICAME website. Have we reached the stage of ICONQUERED? This seems to me (Stig) irrelevant. We are not waging a war. What is important, as far as I can see, is not to conquer but to move on. If ICAME continues to develop, it will have a future as well.

## Notes

1. Originally the Brown Corpus was called in full ‘A Standard Sample of Present-day Edited American English for use with digital computers’.
2. This Windermere speech is printed in full in *ICAME News* 10, pp. 5–7. Nelson Francis’s original handwritten version of the speech is reproduced on the ICAME website.
3. See Francis (1982) in the first ICAME book-length collection of papers, Johansson (1982).
4. This was not the first computer collection of British English texts to get under way. In the 1960s John Sinclair (at first with Michael Halliday at Edinburgh, and then at Birmingham after John’s move there) had put together a collection of over 100,000 words of transcribed spoken English; in the early seventies Jan Aarts (Nijmegen) computerized about 100,000 words of English texts, and Leopold Engels (Leuven) compiled a corpus of British English drama scripts. However, none of these aimed at or achieved anything like the carefully structured, sampled, edited form of the Brown Corpus. Also unique to the Brown Corpus at that stage, and copied by LOB, was the concept of a corpus as a *general resource*, to be copied and distributed for the use of researchers all over the world.
5. At this point some personal acknowledgements should be given, although we can only mention the most important contributors to our work in 1970–77. In the early years at Lancaster, my colleague Norman Fairclough took a role as Assistant Director of CAMET. In building the corpus, the research fellows Rosemary Leonard and Helen Goodluck at Lancaster took a key part, as did Mette-Cathrine Jahr in Oslo. Fanny Leech did some invaluable unpaid work progressing the computerization of the corpus in the bleak period of 1976. At Bergen, the important role of computational support was provided by Knut Hofland, whose expertise was indispensable to ICAME in its early stages, and has remained so up to the present day.
6. On the other hand, if we had relied on e-mail, it is quite likely that the correspondence would not have survived. As part of a small exhibition on the history of ICAME, it is planned to display scanned copies of a range of documents, including some of this correspondence, at the May 2009 meeting of ICAME at Lancaster (ICAME30).
7. This sounds grand, but it should be explained that the delegation consisted of two people, Christine Nuttall of the British Council and myself.
8. At that time the term ‘POS-tagging’ (part-of-speech tagging) was not used. Instead, the talk was of ‘word tagging’, ‘grammatical tagging’, or simply ‘tagging’.

9. For information on the current services through ICAME, see the ICAME website: <http://icame.uib.no/>.
10. Later ICAME bibliographies were to follow, thanks to the dedicated work of Bengt Altenberg; see the ICAME website. See also Altenberg (1991).
11. An earlier list summarising the circulation of the newsletter is found in *ICAME News* 8 (1984), p. 3.
12. The distribution of the newsletter/journal was handled until 2001 by the Bergen Centre. In the period 2001–2005 Eric Atwell, University of Leeds, acted as distributor. The distribution has now been taken over by Paul Rayson at the University of Lancaster.
13. Jostein Hauge was present during part of the founders' meeting in 1977. As he did not stay until the end, he could not sign the document reproduced in Figure 1 above.
14. The title of a paper by Svartvik (1996).
15. Jan Aarts is also part of the editorial team and has played an important role for many years as review editor.

## References

- Altenberg, Bengt. 1991. A bibliography of publications relating to English computer corpora. In S. Johansson and A-B. Stenström (eds.). *English computer corpora. Selected papers and research guide*, 355–396. Berlin and New York: Mouton de Gruyter.
- Francis, W. Nelson. 1982. Problems of assembling and computerizing large corpora. In S. Johansson (ed.). *Computer corpora in English language research*, 7–24. Bergen: Norwegian Computing Centre for the Humanities.
- ICAME News*. Newsletter of the International Computer Archive of Modern English. Bergen: Norwegian Computing Centre for the Humanities.
- Johansson, Stig (ed.). 1982. *Computer corpora in English language research*. Bergen: Norwegian Computing Centre for the Humanities.
- Johansson, Stig, Geoffrey Leech and Helen Goodluck. 1978. *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computers*. Oslo: Oslo University, Department of English.
- Svartvik, Jan. 1996. Corpora are becoming mainstream. In J. Thomas and M. Short (eds.). *Using corpora for language research. Studies in the honour of Geoffrey Leech*, 3–14. London and New York: Longman.