The Organizing of Scientific Fields: The Case of Corpus Linguistics

L A R S E N G W A L L and T I N A H E D M O

Department of Business Studies, Box 513, Uppsala University, Sweden.
E-mail: Lars.Engwall@fek.uu.se

This paper focuses on the processes through which scientific fields are organized over time. It is argued that new approaches in scientific work are hampered by authority structures within national systems for research and established approaches within disciplines, but that these obstacles can be overcome by means of external funding, particularly through new funding sources, as well as the international developments of an innovation. As far as the latter are concerned, they are expected to first lead to informal collaboration among scholars. In the passage of time this informal collaboration becomes more and more formalized. In order to analyse such processes the paper presents a model with three phases labelled as creating, gathering and communicating. This model is then used in an empirical study of corpus linguistics, i.e. the systematic analysis of well-defined populations of written and/or spoken language material. It is shown in the paper how corpus linguistics was developed by scientific innovators who were initially questioned. With the passage of time they created a number of international organizations, which have eventually become more and more formalized, many of them publishing their own journals. In this way the paper demonstrates the significance of organizing for the development of scientific fields.

1. Introduction

Scientific activities are characterized by having higher task uncertainty and mutual dependence between actors than most other human activities.[1] There are variations across different scientific fields in both these dimensions. In terms of task uncertainty physicists are at the low end of the scale, while sociologists, and many other social scientists, are at the other end. Similarly, in terms of mutual dependence physicists are in relative agreement and share beliefs regarding theoretical structures and methodological approaches, while social scientists are much more open regarding both research problems and research methods. However, these variations in task uncertainty and mutual dependence are not given; they vary over time. They are therefore particularly interesting to observe as new scientific fields emerge, as such
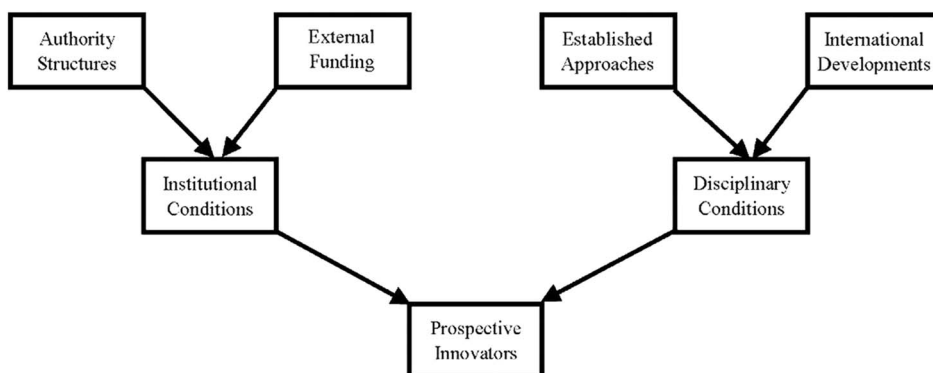
Figure 1. Conditions for prospective innovators.

situations imply that innovators introduce new research approaches with greater uncertainty. In so doing, they try to convince scientific elites and public and private funding agencies that these new tasks are worth pursuing in the hope that they will be acknowledged to the mutual dependence of the scientific community.

Our point of departure for this paper is that there are two types of conditions that are important for prospective innovators (Figure 1): institutional conditions and disciplinary conditions. While the first of these are mainly associated with the organization of academia within specific geographical areas, the latter have to do with the conditions of specific disciplines.

In terms of the institutional conditions (left-hand side of Figure 1) we have identified authority structures and external funding as significant conditions for innovation. The first condition refers to the national regulation of academic work, i.e. rules for the establishment of new institutions and for their governance. In many countries, these rules are subject to political decisions. Governments decide on the structure of academic institutions as well as the rules for employment decisions and resource allocation. In cases with strong authority structures we may expect innovations to be hampered, while the opposite may be true in systems characterized by diversity.

The above means that the control of critical resources by established actors has important implications for innovation. Therefore, the availability of external funding, i.e. that individual scholars rather than formal leaders of organizations are applying for funds, can be expected to counterbalance the effects of hierarchical authority structures, despite the fact that external funding bodies to a considerable extent are also controlled by scientific elites.

While institutional conditions vary among nations, the disciplinary conditions (right-hand side of Figure 1) vary between scientific fields. These conditions, we argue, are determined by the established approaches and international developments. In terms of the former we claim, following Ref. 1, that the lower the task uncertainty and the higher the dependency between researchers, the stronger the resistance towards new theories and new methods will be, and vice versa. However, just as
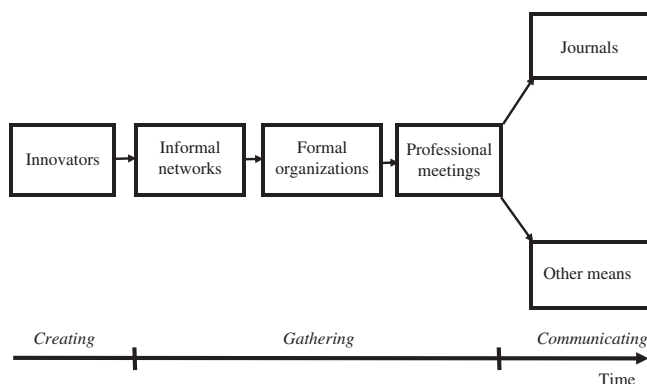
Figure 2. A model for the organizing of scientific fields.

external funding may counterbalance authority structures, international develop-ments of new approaches can fulfil the same function in relation to established approaches. Therefore, it can be argued that the international organizing of a scientific field constitutes a significant part of its development. This means that the dependence between researchers is successively increased over time, leading to a higher integration of the field. This is also the problem we will address in the current paper, i.e. how do scientific fields become organized over time? In dealing with this problem we will use the model for analysis presented in the following section.

2. The Organizing of Scientific Fields

In the introduction we pointed out that scientific innovators can be expected to meet resistance from actors who have the formal power within their institutional system, particularly those affiliated with established approaches in their field of research. A means to overcome such resistance is the establishment of informal networks, particularly in an international context (Figure 2). In order to spread the ideas to new generations these often organize summer schools, which have tended to be very important for the development of new fields.

As new approaches gain ground we can expect the informal networks to turn into formal organizations with statutes for governance, elections, boards, presidents and fees. Since this process is a bottom-up process, it may very well result in several different organizations that support and communicate the new ideas but also compete for prestige. A major task for these organizations will be to gather scholars in the field in workshops, symposia and conferences. With the passage of time these tend to become more and more regular and advanced, turning into significant places for meeting colleagues, for the presentation of ideas and for labour-market talks. The latter is particularly the case in the United States, but as European university systems are increasingly deregulated, conferences tend to become significant for the academic labour markets there as well.

As the organization of a field becomes formally organized, the publication of journals constitutes a further significant step. The conferences produce papers that

tend to be rejected by the established journals, and the scholars in the new field therefore feel that new journals are needed. In view of the large number of new titles appearing, many of these proposals seem to be accepted by the publishing houses. In those cases where they see a potential in the field, the reason for their interest is twofold. First, there are good business opportunities in journals, as profits thus far have tended to be substantial, with high subscription prices for libraries and low costs for editorial screening through the voluntary referee work of academics. Second, particularly in emerging fields it is important for the publishing houses to maintain good relations with the rising stars, who may be the authors of best-selling text-books and but also, together with their colleagues, significant gate-keepers in the selection of literature for their students. As journals become established, the citations game starts, with efforts to raise impact factors, which in turn will attract more and better manuscripts to the journal so that the impact factor can rise even more, and so on.

A further step in the development of a field is the publication of not only journals and textbooks, but other means of communicating, among which electronic discussion forums have become increasingly important. As the field develops, we can expect that prominent scholars in a field are gathered for the publication of handbooks, to which they contribute by writing papers on their expertise. Again, these may become interesting projects for publishing houses on the expectation that the handbooks will be considered a must for university libraries. At the same time they constitute a manifestation of an acknowledgement of the field.

As demonstrated in Figure 2 (bottom) the process can conveniently be divided into three phases of activities: creating (the work of scientific innovators), gathering (informal networks, formal organizations and professional meetings) and communicating (journals and other means of communicating).

The model presented will be used in the following to study empirically the organizing of one specific field: *corpus linguistics*. Following the model we will first, in Section 3, summarize the creating of the field. Then, in Section 4, we will deal with the gathering process, i.e. the development from informal networks to formal organizations with professional meetings. In Section 5 we will subsequently report on studies of communicating efforts, while Section 6 will provide conclusions.

The empirical evidence we are providing below has been collected during the research program 'Re-Structuring Higher Education and Scientific Innovation' (RHESI) within the framework of the European Science Foundation initiative 'Higher Education and Social Change' (EuroHESC). The sources have been (1) publications by corpus linguists and websites and (2) some 50 interviews in France, Germany, the Netherlands, Sweden, Switzerland and the United Kingdom (see Ref. 2).

3. The Creating of Corpus Linguistics

Corpus linguistics can be defined as the systematic analysis of well-defined populations of written and spoken language material. It is a field that has developed particularly in the last 50 years with the advancement of modern computer

technology. However, it is obvious that its roots go further back. Scholars of languages have thus long used corpora for the production of dictionaries, dialect atlases and grammars. As early as the late 19th century the stenography expert Fredrich Wilhelm Kaeding produced a German frequency dictionary.[3] It was followed in the 1920s and 1930s by Vivian Henmon's French frequency word book[4] as well as by the publications of the American and Canadian Committees on Modern Languages.[5–8]

Later on, in the 1950s, the Italian Jesuit Pater Robert Busa made early contributions through his work to provide concordances of the texts of Thomas Aquinas.[9] One of his students, Antonio Zampolli, subsequently became a very active scholar in the field of computational linguistics, not least through the Pisa Summer Schools in the 1970s and the creation of the Pisa Institute of Computational Linguistics (http://www.mt-archive.info/LREC-2004-Zampolli.pdf, see also Ref. 10, p. 35).

Another European pioneer was Bernard Quemada in Besançon, who started his work on computational linguistics in the 1950s. A considerable faculty grant and contacts with the French computer company Bull paved the way for the creation of a laboratory for the study of French vocabulary. Later on he became associated with l'Institut National de la Langue Française (INaLF), an organization founded in 1960 for the development of French lexica, within which he edited 30 volumes of historical French vocabulary.[11] At an early stage he arranged summer schools, which attracted students such as the above-mentioned Antonio Zampolli, and the Manchester scholar Peter Wexler. Among faculty members was the grand old man of French computational linguistics, Charles Muller.[12] Apparently independently of these European-based researchers, Alphonse Juilland (1923–2000) at Stanford published frequency dictionaries of Spanish,[13] Romanian,[14] French,[15] and Italian.[16] Even before that, Randolph Quirk launched the project Survey of English Usage (SEU) at University College London in 1959. In so doing, he turned not only to the collection of written texts but also to spoken English.[17]

Although the scholars mentioned appear to have been frontrunners, Henry Kučera and Nelson Francis, the creators of the Brown corpus at Brown University in Providence, RI, are often mentioned as the pioneers. Their corpus contained around one million words drawn from texts that had been published in the United States during 1961. The original version of the corpus was released in 1964. The lists of the words included in the corpus and the analyses based on them were published in *Computational Analysis of Present-Day American English*;[18] they provided the basis for the first edition of *American Heritage Dictionary* in 1969. The Brown corpus was without doubt an inspiration for many followers in the field of corpus linguistics. One was the CAMET project (Computer Archive of Modern English Texts), created by Geoffrey Leech at Lancaster University, following the same principles as the Brown corpus. It eventually became the Lancaster-Oslo/Bergen (LOB) corpus, completed in 1978 (http://khnt.hit.uib.no/icame/manuals/lob/index.htm).[10]

In Scandinavia, Sture Allén was the first to apply corpus linguistics to Swedish. Through external funding in the mid-1960s he managed to launch a research program that eventually produced a large number of dictionaries.[19–22] In addition, the research

led to the foundation of the Language Bank (*Språkbanken*; see further http://spraakbanken.gu.se/), which was given the task to collect, store, process and provide Swedish texts that could be read electronically. It was established in 1975 as a national centre of computational lexicography. In 1980, Allén was elected one of 18 members of the Swedish Academy, in which he worked as its Permanent Secretary from 1986 to 1999. As a result he was able to link his own work to the long-time project of the Academy to publish an extensive Swedish dictionary.

Afterwards, these achievements of the scientific entrepreneurs of corpus linguists may seem smooth and easy. However, at the time there was considerable evidence of critical attitudes towards the collection of vast databases. The key opponent in this context was the MIT linguist Noam Chomsky, who introduced the idea of the transformational grammar.[23,24] This approach implied that language studies were directed towards the testing of constructions on native speakers instead of the use of corpora. As a consequence, corpus linguists were to a large extent challenged by general linguists, leading to a division between what Charles Filmore[25] has labelled as one between 'armchair linguists' and 'corpus linguists'. No doubt, the corpus linguists felt the negative attitudes. In the words of the Swede Jan Svartvik, who collaborated closely with Randolph Quirk: 'there might have been moments when being named [a corpus linguist, you] felt like discovering your name on the passenger list for the Titanic'.[26] And, as pointed out by Johansson (Ref. 10, p. 33), this 'negative view of corpora found in early generative linguistics persisted in many circles'. Critical attitudes were also found among scholars in literary research. According to Nelson Francis: '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'.[27] These critical attitudes were of course important stimuli to fight back by gathering the corpus linguists, a topic to which we will turn next.

4. The Gathering of Corpus Linguists

4.1. Introduction

It should be clear that this paper focuses on the organizing of scientific fields as a bottom-up process, during which individual scholars develop their contacts and find ways to strengthen their field. This should be distinguished from the top-down organizing constituted by national efforts to document languages. In Section 3 two examples of the latter were mentioned. The first one is the American and Canadian Committees on Modern Languages in the 1920s and 1930s, which had the purpose of improving language acquisition in the United States and Canada. And, the second example is the French l'Institut National de la Langue Française (INaLF) aiming at developing lexica. However, although the mentioned type of top-down efforts have been very important we will leave them for the rest of the analysis.

For the bottom-up processes we should first note that summer schools arranged by the frontrunners have been significant projects for the organizing of the field. Examples mentioned above are the summer schools arranged by the Frenchman Bernard Quemada and the Italian Antonio Zampolli. Another example is the Swede

Table 1. Organizations created in the 1960s

Organization	Acronym	Founded	Area
Association of Computational Linguistics (founded as Association for Machine Translation and Computational Linguistics)	ACL (AMTCL)	1962	North America
International Committee on Computational Linguistics	ICCL	~1965	Transnational
International Computer Archive of Modern and Medieval English	ICAME	1969	Europe

Sources: The web-sites of the organizations and the references provided in the text.

Sture Allén, who in 1972 started Nordic summer schools with another pioneer, Martin Kay − earlier President of the first organization in the field, ACL (see below) − on the faculty.[28] These summer schools were important in spreading the idea of modern corpus linguistics as well as in creating early networks of scholars. Over time, the informal networks increasingly developed into formal organizations. This occurred from the early 1960s onwards. The development first occurred in the United States but was followed by similar projects in Europe. In 2005 even a transnational collaboration between North American and European organizations was formalized.

4.2. Organizations Created in the 1960s

As shown in Table 1, the first formal organization in the field was the Association for Machine Translation and Computational Linguistics (AMTCL), which was founded in the United States in 1962 as an 'international scientific and professional society for people working on problems involving natural language and computation'. In 1968, its name was changed to Association of Computational Linguistics (ACL) to reflect the international character of the organization. This was underlined even more as ACL by the creation in 1982 of a European Chapter (EACL).[29] One of the driving forces behind this step was the perceived need for a specific arena for European researchers to meet. It was also argued as being more appropriate to set up a regional association within ACL than to create a separate and rivalling organization. Today, EACL has developed into being the main professional association for computational linguistics in Europe (www.aclweb.org).

ACL organizes conferences each year, either jointly with its related chapters or alone. In 1989, the organization expanded its activities by establishing the *Data Collection Initiative* (ACL/DCI). This was a response to the increased interest in computational studies for large bodies of text, and was based on the principle that the text corpus should be available for scientific research for a specific fee and without royalties.[30] Two years later, in 1991, ACL also founded the Consortium for Lexical Research (CLR) with grants from an agency of the US Department of Defense, called

the Defense Advanced Research Project Agency (DARPA), at the Computing Research Laboratory in New Mexico to operate as a 'clearinghouse' in the United States and internationally, for samples of lexical data and software (http://aclweb. org/anthology-new/H/H92/H92-1114.pdf). A more recent innovation of ACL is the establishment of 'special interest groups' in areas such as computational linguistics. These provide activities in specific areas in linguistics within ACL's field and related areas by means of workshops, newsletters etc. (www.aclweb.org)

The ACL was followed in the mid-1960s by the International Committee on Computational Linguistics (ICCL). It was set up by David Hays, as a permanent organization exclusively to run international computational linguistics conferences 'but in an original way'.[31] In practice, this meant that the organization should not have a permanent secretariat, subscriptions or funds. The organization started what became one of the international key conferences in computational linguistics, COLING.[32] The conference is arranged every second year and the conference proceedings have been, since 1988, available and distributed through the assistance and cooperation of ACL. The 12th COLING conference was held in Bombay/Mumbai in India in 2012 (http://nlp.shef.ac.uk/iccl/).

The third organization to be created in the 1960s, the International Computer Archive of Modern English (ICAME), provides a prime example of an informal networking that developed into a formal organization, which in course of time has become a central organization for corpus linguists.[33] Its history goes back to 1969 and the University of Lancaster where 'a small group of young and inexperienced academics sat around a table to discuss' how to put Lancaster on the map for research on the English language (Ref. 34, p. 6). The idea was to establish an international organization for archiving, documenting and distributing computer corpus resources, starting with the Brown Corpus, the 'embryonic' Lancaster Corpus and the Survey of English Usage corpus in London. This group of young scholars happened to be those scholars who later became renowned as the frontrunners in modern English corpus linguistics, i.e. Randolph Quirk, Nelson Francis, Geoffrey Leech, Stig Johansson and Jan Svartvik. The founding of the organization was delayed due to difficulties relating to computational inexperience, primitive computing facilities and copyright problems, especially regarding the Lancaster corpus. After 'a flurry of urgent letters' passing among the founding fathers, it was decided to move the 'plan' (including the Lancaster Corpus project) to Oslo where ICAME was formally created in 1977.[34] As a result, ICAME has become a formal organization, with a Constitution, a Chairman and an advisory board with Stig Johansson from the University of Bergen as the first Chairman for almost 20 years (1979–1996).[35] In this way the organization has grown extensively from being a small exclusive 'club' of a small number of researchers to one of the more important players and professional associations in this field, also attracting members from related fields in linguistics. To illustrate, in 1996, 'Medieval' was introduced in the name (keeping the old acronym) in order to recognize the 'flourishing of historical corpus work' inside the organization. This happened as Matti Rissanen, the founder of the Helsinki Corpus, took over as Chairman. However, even earlier the agenda of activities of

Table 2. Organizations created in the 1970s

Organization	Acronym	Founded	Area
Association for Literary and Linguistic Computing	ALLC	1973	Europe
Nordic Association of Linguists	NAL	1976	Europe
Association for Computing and the Humanities	ACH	1978	North America

Sources: The websites of the organizations and the references provided in the text.

ICAME had expanded considerably. In 1979, ICAME started to organize conferences on an annual basis where researchers could meet to coordinate research efforts and to avoid the duplication of research. The first conference was arranged in Bergen to prepare the grammatical tagging of the LOB Corpus, an event gathering 37 members from 10 countries, including the pioneers. Since then, the scope of conferences has broadened considerably in terms of themes and participants. The conference has changed from being simply an event for scholars in English corpus linguistics to one that also engages scholars from related fields (Ref. 33, p. 73).[36] The 35th ICAME conference was held in Nottingham in early May 2014 (see further http://icame.uib.no/).

4.3. Organizations Created in the 1970s

The 1970s saw the creation of another three organizations in the field (Table 2). The first was the Association for Literary and Linguistic Computing (ALLC). It was formally founded in 1973 at a conference at King's College London, but was preceded by two conferences on literary and linguistic computing, the first at the University of Cambridge in 1970, and then at the University of Edinburgh. Founders were the Italian Pater Robert Busa (see above) and Roy Wisbey of King's College, the latter being the first President from 1973 to 1978. Originally, ALLC was directed towards text analysis and language corpora but in the course of time it also included 'history, art history, music, manuscript studies, image processing, and electronic editions'. ALLC has organized conferences annually, first alone and since 1988 jointly with ACH (see below), alternating the venues between Europe and North America. After the creation in 2005 of ADHO (see below), conferences became a joint venture between ALLC, ACH and the Canadian SH-SEMI (see further below). The first joint conference held in Paris in 2006 was followed by annual conferences held alternately on both sides of the Atlantic. For example the 2011 conference was held at Stanford and the 2012 conference in Hamburg (http://www.allc.org/).[37]

The 1970s also saw gathering efforts in geographically restricted areas. Thus, the Nordic Association of Linguists (NAL) was founded in Austin, Texas (!) in 1976 and soon became the main organizational forum in linguistics in the Northern part of Europe. Central to this foundation was the idea of creating a specific society for

Table 3. Organizations created after the 1970s

Organization	Acronym	Founded	Area
European Association for Lexicography	EURALEX	1983	Europe
Canadian Society for Digital Humanities	SDH-SEMI	1986	North America
International Association for Machine Translation	IAMT	1991	Transnational
Linguistic Data Consortium	LDC	1992	North America
European Corpus Initiative Multilingual Corpus	ECI/MCI	1992	Europe
Alliance of Digital Humanities Organizations	ADHO	2005	Transnational

Sources: The websites of the organizations and the references provided in the text.

Nordic linguists. The diffusion and expansion of general linguistics characterizing the Nordic countries in the early 1960s and 1970s was remarkable, and this development required not only some sort of coordination between researchers but also the strengthening of permanent publication channels. As a response, and following a tentative process, NAL was structured as a large organizational platform for Nordic linguists and language scholars as well as linguists outside the Nordic countries studying these languages. From the very start, NAL organized two series of regular international conferences – the International Conference of Nordic and General Linguistics, emphasizing the historical and descriptive study of the Scandinavian languages, and the Scandinavian Conference of Linguistics, being concerned with general linguistic and theoretical issues and a wider range of languages (http://www.uef.fi/nal).[38,39]

The third organization to be created in the 1970s was the Association for Computing and the Humanities (ACH), founded in 1978. It is basically a US organization, which on its website describes itself as 'the major US professional association for computing humanists' and as 'a forum for the research, discussions, and technical explorations'. The leadership of the organization has mainly been American.[40] As mentioned above, ACH has been arranging joint conferences with ALLC since 1988 and within the ADHO collaboration since 2006 (see further www.ach.org).

4.4.   Later Developments

In the following decades the European Association for Lexicography (EURALEX) was established in 1983 (Table 3). This initiative was a response to a conference called LEXeter '83 arranged by Rienhard Hartmann, a well-known lexicographer and applied linguist. EURALEX describes itself as the European-based 'leading professional association for people working in lexicography and related fields' (www.euralex.org), and it acts as an arena for the exchange of ideas for 'lexicographers, reference publishers, corpus linguists, computational linguists, academics working in relevant disciplines, software developers, and anyone with a lively interest in

language' (www.euralex.org). Like the organizations above, EURALEX arranges conferences. They take place every second year, and in July 2014 the organization held its 16th conference, in Bolzano-Bozen, Italy. EURALEX also sponsors smaller events in specific areas in the broader linguistic field (see further www.euralex.org).

Three years after the foundation of EURALEX, in 1986, the gathering efforts were manifested in Canada through the creation of the Canadian Society for Digital Humanities/Société pour l'étude des médias interactifs (SDH-SEMI, http://sdh-semi.org). The purpose of this organization was to 'draw together humanists who are engaged in digital and computer-assisted research, teaching, and creation'. A particular feature of SDH-SEMI is that it is directed towards interaction between anglophone and francophone groups in Canada. As mentioned above SDH-SEMI takes part in the ADHO collaboration for the arrangement of conferences (see further http://sdh-semi.org).

The spread of organizations hosting and distributing computer-based corpora at the international level has continued in the 1990s. One example is the International Association for Machine Translation (IAMT) that was created in 1991 with three chapters: one for the Asia-Pacific Region (AAMT), one for the Americas Region (AMTA) and one for the European Region (EAMT). The purpose of the organization is to organize conferences and workshops. IAMT differs from the other organizations by also having corporate members (www.iamt.org).

The relationship to corporations also applies to the Linguistic Data Consortium (LDC), which was founded the year after IAMT, i.e. in 1992. Its mission is to operate as an open forum for universities, companies and government research laboratories for creating, collecting and distributing speech and text databases, lexicons and other resources for research and development purposes. The LDC is hosted by the University of Pennsylvania and was founded through grants from the US public agency Defense Advanced Research Project Agency (DARPA, www.ldc.upenn.edu). Presently, about 100 companies, universities and government agencies are part of the consortium. It contains 'an indexed collection of Arabic, Chinese and English newswire text, millions of words of English telephone speech from the Switchboard and Fisher collections and the American English Spoken Lexicon, as well as the full text of the Brown corpus'. It can be accessed by members and texts by means of standard browsers (https://online.ldc.upenn.edu/login.html).

In the same year as LDC was created (1992), a similar initiative was taken in Europe through the creation of the *European Corpus Initiative Multilingual Corpus* (ECI/MCI). Its purpose is to 'oversee the acquisition and preparation of a large multilingual corpus […] to be made available in digital form for scientific research at a low cost'. It contains 98 million words of major European languages, Turkish, Japanese, Russian, Chinese, Malay, and more. (http://www.elsnet.org/resources/eciCorpus.html).[41,42]

Yet another step in the gathering of corpus linguists was the creation in 2005 of the umbrella organization – or what Brunsson and Ahrne[43] call the meta-organization – the Alliance of Digital Humanities Organizations (ADHO). The aim of ADHO is to promote and support digital research and teaching in the humanities. Members are

the above-mentioned ACH, ALLC and SDH-SEMI (http://digitalhumanities.org). ADHO organizes and supports an annual conference together with its members (and other organizations), in various constellations within its interest, and a summer school, the Digital Humanities Summer Institute, to open up for discussions and the exchange of knowledge about new computing technologies (http://digitalhumanities.org).

4.5.   Conclusions

The above account has shown how initiatives have been taken on both sides of the Atlantic to create organizations for scholars in the field of corpus linguistics. The development started in the 1960s with Association of Computational Linguistics (ACL) and the International Committee on Computational Linguistics (ICCL) in the United States, which were followed by the European International Computer Archive of Modern and Medieval English (ICAME), the Association for Literary and Linguistic Computing (ALLC) and the Nordic Association of Linguists (NAL) as well as the US-based Association for Computing and the Humanities (ACH). A number of other organizations were created after them. In that development it is particularly worth noting the collaboration across the Atlantic between first ALLC and ACH, and later on SDH-SEMI in an effort to make the conferences more transnational. The same was true as early as the beginning of the 1980s when ACL created a European and a North American chapter and in the early 1990s when the International Association for Machine Translation (IAMT) was founded with its three chapters in three parts of the world. The latter organization also differs from the other organizations in another respect by having corporate members, which is also true for the Linguistic Data Consortium. Together, these observations point to the organizing of scientific field as a gradual process, during which formal organizations are successively created regionally and, over time, broaden their geographical coverage through organizational solutions with chapters or through collaboration. In this process projects for communicating constitute a significant feature, something we will turn to next.

5.   Communicating among Corpus Linguists

5.1.   Journals

According to our reasoning above, the establishing of journals is a natural step following the gathering of the scholars of a new field. Within corpus linguistics this has indeed been the case (Table 4) with the creation of new journals from the mid-1970s and onwards. First out was ACL with its *Computational Linguistics*, which took 13 years to set up after the foundation of the organization.

ACL was followed by ICAME, which, nine years after its foundation launched a publication. This case nicely demonstrates the gradual development of a journal. It started with a newsletter, *ICAME News*, which gave modest information on the ICAME corpora and how they could be reached but also some general information on related projects, meetings and relevant conferences. Over the years, the bulletin grew and came also to include articles, reviews, conference reports, etc. In 1987, it was

Table 4. Significant journals of the field

Journal name	Start	Organization	Time from foundation	Publisher	Impact
Computational Linguistics	1974	ACL (AMTCL)	12	MIT Press	0.940
ICAME Journal	1978	ICAME	9	Lancaster University[a]	n.a.
Nordic Journal of Linguistics	1978	NAL	2	Cambridge University Press	0.318
Literary & Linguistic Computing	1986	ALLC, ACH, SDH-SEMI	13, 8, 0	Oxford University Press	0.717
International Journal of Lexicography	1988	EURALEX	5	Oxford University Press	0.857
International Journal of Corpus Linguistics	1995	–	–	John Benjamins	0.592
Language Resources and Evaluation (Before 2005: Computers and Humanities)	2005		–	Springer	0.659
Corpus Linguistics and Linguistic Theory	2005	–	–	De Gruyter	0.905
Corpora	2006	–	–	Edinburgh University Press	n.a.
Digital Humanities Quarterly	2007	ADHO	2	ADHO	n.a.

Source: Websites of the journals. Impact figures refer to 2012.
Note: [a]Since 2014 ICAME Journal is published by De Gruyter (see http://www.degruyter.com/view/j/icame).

changed into *ICAME Journal*, to reflect the change and content of the publication. Later, all the issues of *ICAME News* and *ICAME Journal* became available online on the website of the organization. (http://icame.uib.no/journal.html).[34] Since 2014 the journal has been published by De Gruyter, a still further step in its institutionalization.

The Nordic linguists were faster than both ACL and ICAME, and launched their journal, the *Nordic Journal of Linguistics*, focusing on theoretical linguists and the languages used in Scandinavia, a mere two years after the foundation of NAL (1978). ALLC, on the other hand, like ACL, needed more than a decade to create *Literary & Linguistic Computing*. This journal is particularly interesting, as today it is a joint publication for ALLC, ACH and SDH-SEMI. This means that the time to a journal was shorter for the collaborators of ALLC, particularly for SDH-SEMI.

The other two journals with links to associations are the EURALEX publication *International Journal of Lexicography,* and the ADHO journal *Digital Humanities Quarterly*. Both appeared in a later part of the process after a relatively short time, which could be a reflection of the increasing pressures to publish (www.euralex.org, http://ijl.oxfordjournals.org/ and www.alc.org).

A comparison of Tables 1–3 and Table 4 reveals that three organizations have not launched a journal. They are ICCL, which decided not to become a formal organization and merely to focus on arranging the COLING conferences, as well as IAMT and LDC, which both have a more commercial orientation through their corporate members.

Needless to say the journals associated with the organizations have certain advantages, since membership usually includes a subscription to the journal of the organization. Therefore, the establishment of other journals may be at a disadvantage. Nevertheless, in the 1990s and the 2000s four new independent journals appeared in the field: *International Journal of Corpus Linguistics*, *Language Resources and Evaluation* (earlier *Computers and Humanities*), *Corpus Linguistics and Linguistic Theory*, and *Corpora*.

It is obvious from Table 4 that UK publishers have taken a particular interest in the publication of journals in the field: Oxford University Press (two titles), Cambridge University Press, Edinburgh University Press and Lancaster University. Another three titles are published by Continental European publishers: the Dutch John Benjamins, and the German De Gruyter and Springer, while only *Computational Linguistics* is put out by the US publisher MIT Press. *Digital Humanities Quarterly* constitutes a special case as it is published by the transnational organization ADHO.

As far as impact is concerned we may note from Table 4 the usual picture: the oldest specialized journal, *Computational Linguistics,* which is published in the United States, has the highest impact factor (0.940), while the others either have lower figures or no such figures available. However, two of the younger journals, published in Europe, have succeeded in attaining impact factors close to that of *Computational Linguistics*: *Corpus Linguistics and Linguistic Theory* (0.905) and *International Journal of Lexicography* (0.857). We can also note that the regionally oriented *Nordic Journal of Linguistics*, not unexpectedly, has the lowest impact factor (0.318) of those that have such a factor available.

Table 5. National origin of editors and editorial board members in 2012

Journal name	Editor(s)	Editorial board		
		North America	Europe	Other
Computational Linguistics (CL)	Other (Australia)	46%	46%	8%
ICAME Journal[a]	Europe (Norway and Sweden)	17%	78%	5%
Nordic Journal of Linguistics (NJL)	Europe (Denmark) and North America	10%	87%	3%
Literary & Linguistic Computing (LLC)	Europe (Belgium)	43%	57%	0%
International Journal of Lexicography (IJLex)	Europe (The Netherland)s	31%	63%	6%
International Journal of Corpus Linguistics (IJCL)	Europe (UK)	18%	64%	18%
Language Resources and Evaluation (Computers and Humanities) (LRE)	Europe (Italy) and North America	38%	44%	18%
Corpus Linguistics and Linguistic Theory (CLLT)	North America	21%	71%	8%
Corpora	Europe (UK)	34%	47%	19%
Digital Humanities Quarterly (DHQ)	North America	62%	38%	0%
Average	60%/30%/10%	34%	57%	9%

Source: Websites of the journals. *Computers and Humanities* is not listed, as it had been transformed into *Language Resources and Evaluation* in 2004.
Note: [a]Regarding the *ICAME Journal*, the members of the ICAME Executive Board have been included in the counts.

In the earlier section we saw how the organizations have aimed at wider geographical coverage. The same is true of the journals. An analysis of the national origin of editors and editorial board members in 2012 (see Table 5) reveals that editors and board members come from North America, Europe as well as from other continents. In terms of editors, seven of the ten journals have at least one European and four have at least one North American editor. The only editorship held outside North America and Europe is for *Computational Linguistics*, for which the editor is located in Australia.

A certain dominance of the Europeans can also be seen for editorial boards. Even in *Computational Linguistics*, published in the US, the Europeans are on a par with the North Americans. The only journal where the North Americans dominate the editorial board is *Digital Humanities Quarterly*, where six out of ten of the editorial board members come from North American institutions. For the rest of the journals the Europeans are in the majority. Among them the *Nordic Journal of Linguistics*, *ICAME Journal* and the *Corpus Linguistics and Linguistic Theory* are particularly

Table 6. Persons with affiliation to more than one of the ten journals in 2012

Name	Affiliation	Journals
Stefan Gries	UC Santa Barbara, USA	*CLLT, Corpora, ICAME*
Michaela Mahlberg	University of Nottingham, UK	*Corpora, ICAME, IJCL*
Svenja Adolphs	University of Nottingham, UK	*Corpora, IJCL*
Tony Berber Sardinha	Catholic University of São Paulo, Brazil	*Corpora, IJCL*
Lynne Bowker	University of Ottawa, Canada	*IJCL, IntJLex*
Ronald Carter	University of Nottingham, UK	*Corpora, IJCL*
Katrin Erk	University of Texas at Austin, USA	*CL, LRE*
Tomaz Erjavec	Institute 'Jozef Stefan', Slovenia	*IJCL, LRE*
Julia Flanders	Brown University, USA	*DHQ, LLC*
Sylviane Granger	Université Catholique de Louvain, Belgium	*Corpora, IJCL*
Patrick Hanks	UWE, Bristol, UK	*IJCL, IntJLex*
Susan Hockey	University College London, UK	*DHQ, LLC*
Michel Hoey	Liverpool University, UK	*Corpora, IJCL*
Frank Keller	University of Edinburgh, UK	*CL, CLLT*
Adam Kilgarriff	Lexicography MasterClass Ltd, UK	*CLLT, IJCL*
Merja Kytö	Uppsala University, Sweden	*ICAME, IJCL*
Willard McCarty	King's College, London, UK	*DHQ, LLC*
Tony McEnery	Lancaster University, UK	*Corpora, IJCL*
Charles Meyer	University of Massachusetts, USA	*Corpora, IJCL*
Joybrato Mukherjee	University of Giessen, Germany	*ICAME, IJCL*
Pam Peters	Macquarie University, Australia	*Corpora, ICAME*
Allen Renear	University of Illinois, USA	*DHQ, LLC*
Ute Römer	Georgia State University, USA	*Corpora, IJCL*
Irma Taavitsainen	University of Helsinki, Finland	*Corpora, ICAME*
Yukio Tono	Tokyo University of Foreign Studies, Japan	*Corpora, IntJLex*
John Unsworth	Brandeis University, USA	*DHQ, LLC*

Source: Websites of the journals.

dominated by Europeans: 87%, 78% and 71%, respectively. Nevertheless, both NJL and CLLT have North American editors.

The total number of persons who are either editors or members of the editorial boards of the ten journals is 267. Of these, 26 persons constitute linking pins by being on the board of more than one journal (Table 6). Two of them − Stefan Gries, UC Santa Barbara, and Michaela Mahlberg, University of Nottingham, both of German origin − even hold three memberships. Again we can observe a stronger representation for the Europeans (58%) and particularly for the United Kingdom (38%). In the latter country, the University of Nottingham with the Centre for Research in Applied Linguistics (http://www.nottingham.ac.uk/cral/index.aspx) is particularly well represented, with three representatives (Svenja Adolphs, Ronald Carter and Michaela Mahlberg).

Our analysis also shows that there are some journals that use the persons in Table 6 more than others. As many as 15 persons are associated with the *International Journal of Corpus Linguistics* (IJCL), 13 with *Corpora* and five each for *Literary & Linguistic*

Computing (LLC) and *Digital Humanities Quarterly* (DHQ). One journal, the *Nordic Journal of Linguistics* (NJL) has no overlap with the other nine journals.

5.2.   Other Means of Communicating

In addition to the journals discussed in the previous section, the organizations in the field also use other means of communicating ideas. The Nordic Association of Linguists, for example, also issued a news bulletin, the *Nordic Linguistic Bulletin* (NLB) for many years. This publication was terminated as a reaction to the increased use of electronic communication. The Bulletin was replaced with an electronic discussion list – the *Nordlingnet* – to provide a forum for discussion and debate among subscribers, i.e. the members of NAL. The ambition was also to use this list as a supplement to the more general and internationally oriented *LINGUIST List* (see below), which is open to all linguistic sub-fields.[38]

Developments similar to those in NAL have occurred in the other organizations. The 1990s was thus characterized by an increasing dispersion of email discussion lists or platforms, providing international arenas for discussion and debates among linguists in general and corpus linguists in particular. These lists are offered by the above organizations and are to a great extent financed through donations from publishing houses and subscribers, i.e. the members of the organizations. The main discussion list is the *CORPORA List*, which was created in the early 1990s by ICAME as a service for spreading and exchanging information and questions from corpus-based linguists and researchers of natural language processing (NLP) (http://icame.uib. no/archives/No_17_ICAME_Journal_index.pdf). Another example, also one of the oldest and most renowned lists for linguists, is the international electronic web-based platform the *LINGUIST List*. The list was founded in the early 1990s by Anthony Aristar, a professor of linguistics, at the University of Western Australia. As early as 1994 there were more than 5000 subscribers, and in 1996, it held its first on-line conference, *Geometric and Thematic Structure in Binding*. In addition to donations from supporting publishers, institutions and subscribers the *LINGUIST List* is supported by grants from the National Science Foundation. The list now has subscribers all over the world, and it operates as an arena for queries and the dissemination of results, discussions and debates, journal table of contents, dissertation abstracts, calls for papers, book and conference announcements, etc. Since 2006, all its operations have been located at Eastern Michigan University (www.linguistlist.org).

Another electronic international discussion list is *HUMANIST*, which was started in 1987 by Willard McCarty, a senior lecturer in humanities computing at King's College London, and under the joint sponsorship of the ACH, the ALLC and the University of Toronto's Centre for Computing in the Humanities. It was created as a forum for discussion and exchange of information among subscribers in issues relating to humanities computing in discipline areas such as linguistics, comparative literature, and philosophy.[44] In 2008 there were 1650 subscribers to the list. Over the past decades, *HUMANIST* has morphed into an online publication as well and is

today allied with ADHO (see above) and published by the Office for Humanities Communication (OHC). *HUMANIST* is also an affiliated publication of the American Council of Learned Societies (ACLS) (http://www.digitalhumanities.org/humanist/McCarty__Report_on_Humanist_2008-abbreviated.pdf).

All these efforts to communicate ideas are of course very important for the development and the integration of the field. Other signs of this are a large number of textbooks introducing corpus linguistics.[33,45–47] Further evidence in the same direction comprises two handbooks that were published some years ago.[48,49]

5.3.   Conclusions

Our analysis of the communicating phase has shown how the majority of the organizations have taken initiatives within a decade of their foundation through first conference proceedings and eventually journals. We have also seen how new journals without links to the organizations have been launched as well as a collaboration involving three organizations around one journal (LLC). Most of the journals in the field have European publishers, the exception being *Computational Linguistics* published by the MIT Press. Of the ten journals we have identified, it is also the most influential.

The European presence in the field is not only true for the publishers, but also for those active in these journals. About 60% of the editors as well as editorial members are European. In *Computational Linguistics*, which is published in the US, the Europeans are on a par with the North Americans. Among the Europeans, scholars from Great Britain play a particularly important role in the field. Overall there appears to be an ambition to connect the journals to institutions on both sides of the Atlantic and to a certain extent to other parts of the world. In addition we have been able to show how the journals are related to each other through joint editorial board members. This has particularly been the case with the *International Journal of Corpus Linguistics* and *Corpora*, which each have more than ten persons associated with another journal in the field.

The integration of the field appears also to be reinforced by various projects for electronic communication and sharing of information. Obviously, present-day linguists have excellent opportunities to access corpora that have been established over time. One important example in this context is the European Corpus Initiative Multilingual Corpus with its 98 million words of major European languages. In addition, the recruitment of new corpus linguists is facilitated through a stream of textbooks as well as handbooks dealing with various aspects of the field.

6.   Conclusions

Focusing on the organization of scientific fields, in this paper we have taken as a point of departure the difficulties faced by new ideas and new approaches through resistance from established authority structures within national systems for research as well as established approaches within disciplines. We have also pointed to the opportunities offered by the existence of external funding and the international

developments of an innovation. Thus, in order to overcome resistance, we have argued, scientific innovators tend to organize themselves internationally, a process that includes three significant activities: creating the field, gathering its members, and communicating the research results. This process implies that individual scholars in various countries in the passage of time gradually come together, first in loose networks and later on in increasingly formalized structures with statues, presidents, regular professional meetings, and publications.

These theoretical arguments have been confronted with empirical evidence in a case study of corpus linguistics, i.e. the systematic analysis of well-defined populations of written and spoken language material. This study has confirmed our arguments. As corpus linguistics developed in the 1950s and the 1960s the scientific entrepreneurs met considerable resistance from general linguists. Nevertheless, corpus linguistics developed in different countries, and eventually scholars in the field came to collaborate in various constellations on both sides of the Atlantic. In addition, there were transnational organizations as well as efforts to develop collaboration between North American and European organizations. As predicted, we have also seen the emergence of a number of journals, particularly in collaboration with European publishers, with a majority of Europeans among the editors and editorial boards. And, this European dominance could also be expected due to greater needs for Europeans, in comparison with their US colleagues, to collaborate across countries.

The empirical study of corpus linguistics thus seems to support our theoretical reasoning regarding the organizing of scientific fields. Nevertheless, further studies of the processes involved in the development of new approaches are required. First of all, we need to go further into the case of corpus linguistics in terms of more extensive studies of the creating, gathering and communicating phases of the field. Here, additional studies of scientific innovators are required regarding the frontrunners, those taking initiatives for organizations and journals as well as the development over time of statues, presidents, board members, etc. A particularly interesting feature concerns the differences among countries in terms of the involvement of their scholars in process. A recent comparison of the adoption of corpus linguistics in Germany, the Netherlands, Sweden and Switzerland[2] thus shows that Germans and Swedes were much earlier than their colleagues in the other two countries, and particularly in relation to those in Switzerland.

Second, it is of course important to perform studies of other fields to see whether similar patterns can be observed there. As a matter of fact, we know from studies of the management field that similar processes have occurred there. In the 1970s a number of European associations were created, first informally followed by increasing formalization and the creation of journals (Ref. 50, pp. 172 and 210). For example, the European Group of Organization Studies (EGOS) was created as a loose network in 1973 with no president, no board and no fees. It did not become a formal organization until 1998 when it was registered as an association in Brussels, thereby having to develop the statues and, among other things, having to elect a board.[51]

Finally, it is worth noting that the organizing of scientific fields is part of more general processes in modern society. We can note that these are often the result of initiatives by institutional entrepreneurs who see the need for more coordination

within a field but also the opportunities of going together in order to promote certain ideas or to raise the prestige of members. Sometimes, such initiatives are also the results of individual ambitions to create personal platforms. However, irrespective of the motives, organizing appears to be here to stay and to flourish.

Acknowledgements

References and Notes

1. R. Whitley (1984) *The Intellectual and Social Organization of the Sciences* (Oxford: Oxford University Press).
2. L. Engwall, E. Aljets, T. Hedmo and R. Ramuz (2014) Computer corpus linguistics: An innovation in the humanities. *Research in the Sociology of Organizations*, **42**, pp. 329–363.
3. F.W. Kaeding (1897-1898) *Häufigkeitswörterbuch der deutschen Sprache 1-2* (Steiglitz bei Berlin: Selbstverlag des Herausgebers).
4. V.A.C. Henmon (1924) A French word book based on a count of 400 000 running words *Bureau of Educational Research Bulletins 3* (Madison: University of Wisconsin).
5. G.E. Vander Beke (1929) *French Word Book*. Publications of the American and Canadian Committees on Modern Languages 15 (New York: Macmillan).
6. M.A. Buchanan (1931) *A Graded Spanish Word Book*. Publications of the American and Canadian Committees on Modern Languages 3 (Toronto: University of Toronto Press).
7. F.D. Cheydleur (1934) *French Idiom List Based on a Count of 1,183,000 Running Words*. Publications of the American and Canadian Committees on Modern Languages 16 (New York: Macmillan).
8. B.Q. Morgan (1933) *German Frequency Word Book,* based on Kaeding's Häufigkeitswörterbuch der deutschen Sprache. Arranged and edited by B.Q. Morgan. Publications of the American and Canadian Committees on Modern Languages 9 (New York: Macmillan).
9. See, for example, R. Busa (1951) Sancti Thomae Aquinatis hymnorum ritualium varia specimina concordantiarum: Primo saggio di indici di parole automaticamente composti e stampati. *Archivum philosophicum Aloisianum* Ser, 2; 7. Milan.
10. See, S. Johansson (2008) Some aspects of the development of corpus linguistics in the 1970s and 1980s. In: A. Lüdeling and M. Kytö (eds), *Corpus Linguistics: An International Handbook* (Berlin: Mouton de Gruyter), pp. 33–53.
11. B. Quemada (ed.) (1959-1993) *Matériaux pour l'histoire du vocabulaire français: datations et documents lexicographiques*, Part 1-30 (Besançon: Centre d'étude du vocabulaire français).
12. Cf. for example, C. Muller (1967) *Étude de statistique lexicale: le vocabulaire du théâtre de Pierre Corneille*. Paris: Larousse (Diss. Université de Strasbourg);

C. Muller (1968) *Initiation à la statistique linguistique* (Paris: Larousse); C. Muller (1979) *Langue française et linguistique quantitative: recueil d'articles* (Geneva: Slatkine). For Muller's last Festschrift, in celebration of his centenary, see C. Delcourt and M. Hug (eds) (2009) *Mélanges offerts à Charles Muller pour son centième anniversaire (22 septembre 2009)* (Paris: CILF).

13. A. Juilland and E. Chang-Rodriguez (1964) *Frequency Dictionary of Spanish Words*. The Romance Languages and their Structures. First series, S 1 (The Hague: Mouton).

14. A. Juilland, P.M.H Edwards and I. Juilland (1965) *Frequency Dictionary of Rumanian Words*, The Romance Languages and their Structures. First series, R1 (The Hague: Mouton).

15. A. Juilland, D. Brodin and C. Davidovitch (1970) *Frequency Dictionary of French Words*. The Romance Languages and their Structures. First series, F1 (The Hague: Mouton).

16. A. Juilland, V.P. Traversa and A. Beltramo (1973) *Frequency Dictionary of Italian Words*. The Romance Languages and their Structures. First series, I1 (The Hague: Mouton).

17. R. Quirk and J. Svartvik (1978) *A Corpus of Modern English* (Lund: Lund University).

18. H. Kuçera and N. Francis (1967) *Computational Analysis of Present-day American English* (Providence, RI: Providence University Press).

19. S. Allén (1970-1980) *Nusvensk frekvensordbok baserad på tidningstext: Frequency Dictionary of Present-day Swedish based on Newspaper Material*. Data linguistica 1-4 (Stockholm: Almqvist & Wiksell International).

20. S. Berg (1978) *Olika lika ord: svenskt homograflexikon (On Different Words: Swedish dictionary of homographs)*, Data linguistica 12 (Stockholm: Almqvist & Wiksell International).

21. S. Allén, Å. Abelin and Others (1986) *Svensk ordbok (Swedish Dictionary)* (Solna: Esselte Studium).

22. Already in 1962 Allén had presented a concordance of a 17th century text. Ten years later, in 1972, he was appointed to a chair in computational linguistics, which is considered to have been the first of its kind worldwide. (Interview by Lars Engwall with Sture Allén, 17 November 2011).

23. N.A. Chomsky (1957) *Syntactic Structures* (New York: Mouton).

24. N.A. Chomsky (1965) *Aspects of the Theory of Syntax* (Cambridge, MA: The MIT Press).

25. C. Filmore (1992) 'Corpus linguistics' or 'computer-aided armchair linguistics'. In *Directions in Corpus Linguistics*. Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991, ed. Jan Svartvik. (Berlin: Mouton de Gruyter), pp. 35–60, p. 35.

26. J. Svartvik (2007) Corpus linguistics 25+ years on. In: R. Fachinetti, (ed.), *Corpus Linguistics 25 Years On* (Amsterdam: Editions Rodopi), pp. 11–25, p. 11.

27. N. Francis (1984) Dinner speech. *ICAME News* 10: 5–7.

28. Interview by Lars Engwall with Sture Allén, 17 November 2011.

29. In 1997 a North American Chapter (NAACL) including members from Central and South America was created.

30. M. Liberman (1989) Text on tap: the ACL/DCI, speech and natural language: Proceedings of a workshop held at Philadelphia, Pennsylvania, 21–23 February 1989, H89-2024 (http://acl.ldc.upenn.edu/H/H89/H89-2024.pdf).

31. Hays was the founding chairman of the then newly formed linguistics department at the State University of New York at Buffalo and professor of linguistics, computer science, information and library studies.

32. In addition COLING describes itself as follows: 'COLING has always been distinguished by pleasant venues and atmosphere, rather than by the clinical efficiency of an airport conference hotel: COLINGs are simply nice conferences to be at. They have also striven for inclusiveness, both geographical, when the world was more harshly divided than it is now, and theoretical, in that COLING has been less prone to mood swings of theory than societies that are run in different, and more conventional, ways' (http://nlp.shef.ac.uk/iccl/nature.html).

33. T. McEnery and A. Hardie (2012) *Corpus Linguistics. Method, Theory and Practice* (Cambridge: Cambridge University Press).

34. G. Leech and S. Johansson (2009) The coming of ICAME. *ICAME Journal*, **33**, pp. 5–11.

35. According to Geoffrey Leech, the main reason for the formalisation of ICAME was to solve copyright problems with the LOB corpus (Interview by Lars Engwall with Geoffrey Leech, 9 May 2013).

36. S. Johansson (2009) The early history of ICAME. *ICAME Journal*, **33**, pp. 12–20.

37. S. Hockey (2004) The history of humanities computing. In: S. Schreibman, R. Siemens and J. Unsworth (eds), *A Companion to Digital Humanities* (Oxford: Blackwell), Chapter 1.

38. S. Eliasson (2010) The Nordic Association of Linguists: The preparatory phase and the first thirty years (1977–2006). In: H. Götzsche, (ed.) *Memory, Mind and Language* (Newcastle upon Tyne: Cambridge Scholars Publishing). 4–54.

39. An even earlier effort in Scandinavia was *Svenskans beskrivning*, which grew out of a conference at Stockholm University in 1963 focusing on phonetics, albeit with a light touch of structuralism. Since then conferences have been arranged approximately every 18 months with increasingly standardized proceedings. These Nordic developments were also stimulated by summer schools from the 1960s and onwards. In addition, numerous additional Nordic conferences were started during this period, such as the Nordic Conference on Computational Linguistics (NODALIDA) starting in Gothenburg in 1977. See E. Hovdhaugen, F. Karlsson, C. Henriksenl and B. Sigurd (2000) *The History of Linguistics in the Nordic Countries* (Helsinki: Societas Scientiarum Fennica); F. Karlsson (2008) Svenskans beskrivning 1-29. In: C. Falk, A. Nord and R. Palm (eds), *Svenskans beskrivning 30* (Stockholms Universitet: Institutionen för nordiska språk).

40. The first President was Joe Raben at Queens College, City University of New York, who was followed by other US professors until 2004. That year a European, Lorna Hughes from the University of Wales, was elected President. However, as she stepped down at the end of 2007 she was succeeded by US scholars, first Julia Flanders from Brown, and then by Bethany Nowviskie from the University of Virginia.

41. K.W. Church and R.L. Mercer (1993) Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, **19**, pp. 1–24.

42. Since 1999, conferences have been arranged about every 18 months by the American Association of Corpus Linguists (AACL). However as pointed out to Merja Kytö by Randi Reppen, who organized the 2014 conference in Flagstaff, AACL 'is not really an organization' (Personal communicationo).

43. N. Brunsson and G. Ahrne (2008) *Meta-organizations* (Cheltenham: Edgar Elgar).
44. W. McCarty (1992) HUMANIST: Lessons from a global electronic seminar. *Computers and the Humanities*, **26**, pp. 205–222.
45. C.F. Meyer (2002) *English Corpus Linguistics: An Introduction* (Cambridge: Cambridge University Press).
46. M.A.K. Halliday (2004) *Lexicology and Corpus Linguistics: An Introduction* (London: Continuum).
47. W. Teubert and A. Čermáková (2007) *Corpus Linguistics: A Short Introduction* (London: Continuum).
48. A. Lüdeling and M. Kytö (eds) (2008-2009) *Corpus Linguistics: An International Handbook*, Vols 1-2 (Berlin: Mouton de Gruyter).
49. A. O'Keeffe and M. McCarthy (eds) (2010) *The Routledge Handbook of Corpus Linguistics* (Abingdon: Routledge).
50. L. Engwall (2009) *Mercury meets Minerva* (Stockholm: the Economic Research Institute of the Stockholm School of Economics) (Extended second edition of *Mercury Meets Minerva*, Oxford: Pergamon Press, 1992).
51. Through an initiative of the publisher De Gruyter the journal Organization Studies was launched as early as 1975 (http://www.egosnet.org/egos/about_egos/egos_history_short_overview).

Electronic Sources and Web-pages (accessed in May 2012)

http://aclweb.org/anthology-new/H/H92/H92-1114.pdf
http://digitalhumanities.org
http://icame.uib.no/
http://icame.uib.no/archives/No_17_ICAME_Journal_index.pdf
http://icame.uib.no/journal.html
http://ijl.oxfordjournals.org/
http://khnt.hit.uib.no/icame/manuals/lob/index.htm
http://nlp.shef.ac.uk/iccl/
http://sdh-semi.org
http://spraakbanken.gu.se/
http://www.ach.org
http://www.aclweb.org
http://www.allc.org/
http://www.digitalhumanities.org/humanist/McCarty__Report_on_Humanist_2008-abbreviated.pdf
http://www.eamt.org
http://www.egosnet.org/egos/about_egos/egos_history_short_overview, 2012-05-31
http://www.elsnet.org/resources/eciCorpus.html
http://www.euralex.org
http://www.iamt.org
http://www.ldc.upenn.edu
http://www.linguistlist.org
http://www.nottingham.ac.uk/cral/index.aspx
http://www.mt-archive.info/LREC-2004-Zampolli.pdf
http://www.uef.fi/nal
https://online.ldc.upenn.edu/login.html

About the Author

Lars Engwall is Professor Emeritus of Management at Uppsala University, Sweden, and a member of the Academia Europaea HERCulES group. His research has been directed towards the production and diffusion of management ideas, particularly in media companies, banks and academic institutions.

Tina Hedmo is Lecturer in the Department of Business Studies at Uppsala University, Sweden. Her research has been directed towards academic systems, particularly quality control through accreditation.