

Historical Text Mining Workshop

Thursday 20th and Friday 21st July 2006,
Infolab21, Lancaster University, UK.

Organisers:

Paul Rayson
(Lancaster University)

Dawn Archer
(University of Central Lancashire)

<http://ucrel.lancs.ac.uk/events/htm06/>



Sponsored by the AHRC ICT Methods Network

Workshop programme

Thursday 20th July	
10:00	Registration and coffee
10:20	Welcome and introductions
Morning Session: What's possible with modern data?	
10:30	"Historical Text Mining", Historical "Text Mining" and "Historical Text" Mining: Challenges and Opportunities Robert Sanderson (University of Liverpool)
11:00	Introduction to GATE (General Architecture for Text Engineering) Wim Peters (University of Sheffield)
11:30	Introduction to WordSmith Mike Scott (Liverpool University)
12:00	Tutorial: Corpus annotation and Wmatrix Paul Rayson (Lancaster University)
12:30	Lunch
Afternoon session: Problems re-applying such tools to historical data	
13:30	Search methods for documents in non-standard spelling Thomas Pilz and Andrea Ernst-Gerlach (Universität Duisburg-Essen)
14:00	The Potential of the Historical Thesaurus of English Christian Kay (University of Glasgow)
14:30	Discussion
15:00	Coffee/tea break
15:30	The CEEC corpora and their external databases Samuli Kaislaniemi (University of Helsinki)
15:50	Exploring Speech-related Early Modern English Texts: Lexical Bundles Re-visited Jonathan Culpeper (Lancaster University) and Merja Kytö (Uppsala University)
16:30	Introducing nora: a text-mining tool for literary scholars Tom Horton (University of Virginia)
17:00	Closing discussion

Friday 21st July	
Morning Session: Possible solutions to the problems identified on day 1	
9:30	Teaching a computer to read Shakespeare: the problem of spelling variation and Tutorial on the variant detector tool (VARD) Dawn Archer (University of Central Lancashire) and Paul Rayson (Lancaster University)
10:15	The advantages of using relational databases for large historical corpora Mark Davies (Brigham Young University)
11:15	Coffee/tea break
11:30	Discussion
12:00	Managing Momus: following the fortuna [sic] and frequency of a trope in Early English Books Online Stephen Pumfrey (Lancaster University)
12:15	Digitisation of historical texts at ProQuest and ways of accessing variant word forms Tristan Wilson (ProQuest Information and Learning)
12:30	Lessons learnt from transcribing and tagging the Newcastle Electronic Corpus of Tyneside English Joan Beal (University of Sheffield) and Nick Smith (Lancaster University)
13:00	Lunch
13:30	Visit to the Rare Book Archive in Lancaster University Library to view the Hesketh Collection Participants who have registered in advance of the workshop will need to bring photo-ID.
Afternoon session: Final round-up	
14:00	Software demonstrations and small-group discussion
14:30	Nineteenth Century Serials Edition Project Suzanne Paylor, Jim Mussell (Birkbeck College)
14:45	LICHEN: The Linguistic and Cultural Heritage Electronic Network Lisa Lena Opas-Hänninen (University of Oulu)
15:00	Discussion: Software requirements
15:15	Coffee/tea break
15:45	Round-up discussion: where to next?
16:30	Close

Abstracts for presentations

"Historical Text Mining", Historical "Text Mining" and "Historical Text Mining: Challenges and Opportunities

Robert Sanderson (University of Liverpool)

The title of the workshop is itself a perfect example of the challenges of text mining, as all three groupings of the words are equally valid: the application of text mining to modern treatises about the past; the history of text mining; or the mining of ancient texts, be they fact, fiction, poetry or prose. This presentation will try to cover all three at a high level, suggesting some definitions, challenges and opportunities for meaningful interdisciplinary collaboration, for subsequent presentations to address in more detail.

Dr Robert Sanderson is a lecturer in Computer Science at the University of Liverpool, teaching Data Mining and Information Systems. He completed his PhD in 2003, comprising an electronic edition of Book 1 of the medieval historian Froissart's "Chroniques", merging information science, medieval history and language. Since 2000, he has been working with the Cheshire Project, jointly at UC Berkeley and Liverpool, a partner in the National Centre for Text Mining to provide a standards-based, grid-capable information framework.

Introduction to GATE (General Architecture for Text Engineering)

Wim Peters (University of Sheffield)

In this presentation I will provide an introduction to the language engineering platform GATE (Generalised Architecture for Text Engineering; <http://www.gate.ac.uk>). GATE has been under development for more than 10 years, and is used by thousands of people around the world within research areas such as linguistics, human language technology and semantic web. GATE comes with a user-friendly graphical interface, and enables the user to

- load their documents from a variety of formats, from Ascii to XML;
- select existing language analysis tools, for instance a part of speech tagger or a morphological analyser.

Various modules are freely available for the analysis of a number of languages, although English is by far most widely supported. The analysis tools that are currently available in GATE allow the user to:

- manually/automatically create annotations for text elements (words, phrases, paragraphs...)
- integrate existing tools into the GATE architecture and run them;
- create new tools in order to create self-made annotations
- evaluate the output
- export their results in XML

These functionalities for the analysis of synchronic language use are useful for any diachronic linguistic applications. Proof of concept has already been given by a number of projects that work on historical aspects of language: Perseus, ETCSL and The Old Bailey. If time permits, I will give a live demonstration to illustrate some of the functionalities listed above.

Search methods for documents in non-standard spelling

Thomas Pilz and Andrea Ernst-Gerlach (Universität Duisburg-Essen)

The interdisciplinary project 'Rule-based search in text databases with non-standard orthography (RSNSR)' develops methods to ease the often time consuming work with historical documents. Our goal is to provide expert users as well as interested amateurs with an online search engine that is able to reliably deal with the problems that arise when working with non-standard texts, i.e. historical spellings, faulty character recognition or transcription and obsolete characters. Therefore, we are working on manually as well as automatically derived rule sets and distance metrics and their topographical and diachronic distribution.

The Potential of the Historical Thesaurus of English

Christian Kay (University of Glasgow)

The Historical Thesaurus of English (HT)¹ is a semantic index to the Oxford English Dictionary (OED)² supplemented by Old English materials also published separately in A Thesaurus of Old English (TOE)³. Word senses are organised within 26 major semantic fields in a hierarchy of categories and subcategories, with up to fourteen levels of delicacy. The material is held in a database and first steps towards internet publication have been taken by an AHRC-ICT Strategy Project creating searches for use in a range of humanities disciplines⁴. Although initially of interest to historians of the English Language, the study of vocabulary can also reveal a good deal of social and cultural information. In addition to a browsing facility, searches are currently offered on synonyms, affixes, style labels, parts of speech and dates of use. It will be argued that the project also has potential for text mining and for probability-based disambiguation of polysemous words⁵.

¹ <http://www.arts.gla.ac.uk/sesll/englang/thesaur/homepage.htm>

² *The Oxford English Dictionary*. 1884-1933 ed. by Sir James A. H. Murray, Henry Bradley, Sir William A. Craigie & Charles T. Onions; *Supplement*, 1972-1986 ed. by Robert W. Burchfield; 2nd edn, 1989, ed. by John A. Simpson & Edmund S. C. Weiner; *Additions Series*, 1993-1997, ed. by John A. Simpson, Edmund S. C. Weiner & Michael Proffitt; 3rd edn (in progress) *OED Online*, March 2000-, ed. by John A. Simpson. Oxford: Oxford University Press.

³ Jane Roberts & Christian Kay with Lynne Grundy, *A Thesaurus of Old English*, King's College London Medieval Studies XI, 1995, 2 vols. Second impression, Amsterdam: Rodopi, 2000.

An electronic version, supported by British Academy LRG-37362, can be seen at <http://libra.englang.arts.gla.ac.uk/oethesaurus/>

⁴ Jeremy Smith, Simon Horobin and Christian Kay, Lexical Searches for the Arts and Humanities, AR112456.

⁵ For an overview, see Wilks, Yorick A., Slator, Brian M., and Guthrie, Louise M., *Electric Words: Dictionaries, Computers and Meanings* (Cambridge, MA, and London: MIT Press, 1996). More recent work, including the MALT project (Mappings, Agglomerations and Lexical Tuning), is described at <http://www.dcs.shef.ac.uk/~yorick/>

The CEEC corpora and their external databases

Samuli Kaislaniemi (University of Helsinki)

The *Corpus of Early English Correspondence* (CEEC) is a diachronic corpus of personal letters designed for historical sociolinguistics, compiled at the University of Helsinki since 1994. It spans the years 1402-1800, and contains about 5.3 million words.

The CEEC consists of five subcorpora of texts, and two databases of extralinguistic socioregional information: the first database contains information on each letter, the second detailed information on each writer. In the early stages of compilation of the CEEC, it was felt that including this information in the text header of each letter in TEI- or COCOA-coding did not really fit the design of the corpus. Extensive headers were also felt to present further, technical, problems.

Seeing that such problems can be overcome with modern corpus and database tools, the CEEC team is now working on digitizing the letter database, and creating a relational database architecture which will enable free searches of the corpus texts according to the desired combination of socioregional variables.

At present, two of the CEEC subcorpora have been released through the Oxford Text Archive: the 450,000-word *CEEC Sampler* (CEECS) in 1998, and the 2.2m-word parsed and POS-tagged *Parsed CEEC* (PCEEC) in 2006.

References

- CEECS = *The Corpus of Early English Correspondence Sampler*. 1998. Compiled by the CEEC Project Team. Helsinki: University of Helsinki. <<http://khnt.hit.uib.no/icame/manuals/ceecs>>.
- PCEEC = *The Parsed Corpus of Early English Correspondence*. 2006. Annotated by Ann Taylor, Arja Nurmi, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. <www.eng.helsinki.fi/varieng/team2/1_2_4_projects.htm>
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Longman.
- Raumolin-Brunberg, Helena & Terttu Nevalainen. 2005. "Historical sociolinguistics: The *Corpus of Early English Correspondence*". *Models and Methods in the Handling of Unconventional Digital Corpora* vol. 2: *Diachronic Corpora* ed. by J. C. Beal, K. Corrigan & H. Moisl. Palgrave.

Exploring Speech-related Early Modern English Texts: Lexical Bundles Re-visited

Jonathan Culpeper
Department of Linguistics and English Language
Lancaster University

Merja Kytö
Department of English
Uppsala University

The study of 'fixed' expressions is receiving increasing attention within the corpus linguistics framework. Most studies have so far focused on present-day English, spoken (cf., e.g., Aijmer 1996, Hudson 1998) and written (cf., e.g., Moon 1998). Much less work has been done on linguistic fixedness in early English, despite the fact that the role of conventionalised expressions in the history of discursal phenomena is becoming a popular area of study (e.g. Kopytko 1995; Nevalainen and Raumolin-Brunberg 1995; Nevala 2004; Traugott 2001: chapter 4).

In an earlier study (Culpeper and Kytö 2002), we investigated the role played by recurrent word-combinations in a pilot version of the *Corpus of English Dialogues, 1560-1760*. Biber et al. (1999) refer to such recurrent expressions as 'lexical bundles'—computationally derived groups of words, where each word has a particular frequency of co-occurrence with other words in the group. There is no requirement that the bundles are structurally or semantically complete, and in fact they are usually not. In common with other studies (e.g. Altenberg 1998), we will focus on lexical bundles which consist of at least three words. Biber et al. (1999: 992) suggest that three-word bundles can be considered as a kind of extended collocational association. Once we derived our lexical bundles, we classified them according to function. Our functional classification was broadly based on Halliday's (e.g. 1994) three functional components of language: ideational, interpersonal and textual (see also Moon 1998).

In our present paper we will pursue this topic further, addressing the following research questions: (i) what were the recurrent types of lexical bundles characteristic of authentic and constructed dialogue of the period? (ii) how do they relate to particular functions; and (iii) how do they compare with those found in Present-day spoken English? Our study now has the benefit of a much more extensive collection of early texts drawn from the *Corpus of English Dialogues*. Furthermore, we will make comparisons with data drawn from a comparative corpus of present-day trial proceedings and play-texts. Methodologically, we have also addressed the issue of spelling variation in our early data.

Our study will contribute to our knowledge of the 'spoken' language of the past, the much recognized locus of most linguistic change. It will also illuminate the present through the past and offer a starting-point for further work in the field of variationist study and historical pragmatics.

References

- Aijmer, K. 1996. *Conversational Routines in English. Convention and Creativity*. London & New York: Longman.
- Altenberg, Bengt. 1998. "On the phraseology of spoken English: The evidence of recurrent word-clusters". *Phraseology. Theory, Analysis, and Applications*, ed. by A.P. Cowie, 101–122. Oxford: Clarendon Press.

Abstracts for presentations

- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.
- Culpeper, J. and M. Kytö. 2002. "Lexical bundles in Early Modern English dialogues: A window into the speech-related language of the past", in: T. Fanego, B. Méndez-Naya, and E. Seoane (eds.), *Sounds, Words, Texts and Change. Selected Papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*. Amsterdam/Philadelphia: John Benjamins, 45–63.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Hudson, J. 1998. *Perspectives on Fixedness, Applied and Theoretical*. Lund: Lund UP.
- Jucker, A. H. (ed.). 1995. *Historical Pragmatics. Pragmatic Developments in the History of English*. Amsterdam/Philadelphia: John Benjamins.
- Kopytko, R. 1995. "Linguistic politeness strategies in Shakespeare's plays", in: A. H. Jucker (ed.), 515–540.
- Moon, R. 1998. *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.
- Nevala, M. 2004. *Address in Early English Correspondence: Its Forms and Socio-Pragmatic Functions*. Helsinki: Société Néophilologique.
- Nevalainen, T. and H. Raumolin-Brunberg. 1995. "Constraints on politeness. The pragmatics of address formulae in early English correspondence", in: A. H. Jucker (ed.), 541–601.
- Traugott, Elizabeth Closs. (2001) *Regularity in Semantic Change*. New York: Cambridge University Press.

Introducing nora: a text-mining tool for literary scholars

Thomas B. Horton (University of Virginia, Department of Computer Science)

The high-level goal of the *nora project* (<http://www.noraproject.org/>) is to produce software for discovering, visualizing, and exploring significant patterns across large collections of full-text humanities resources in existing digital libraries. Our initial focus is to apply text-mining to explore problems of interest to literary scholars studying 19th century American literature. Funded by the Andrew W. Mellon Foundation, the project includes members from five universities with a wide-range of backgrounds, from the humanities to visual design to information science and computer science.

In its first 18 months, efforts within the project have focused on several initial problems, described below.

a) Developing a software architecture to support text-mining. A system called *Tamarind* was developed to process XML files, extract features, and store results in a database system. A data-mining system developed by the NCSA called *D2K* (Data to Knowledge, <http://alg.ncsa.uiuc.edu/do/tools/d2k>) gets data from Tamarind's data-stores for a given document collection and then carries out certain text-mining classification algorithms. D2K's Web Services component has been used to allow end-user software applications to invoke text-mining tasks on a server and get results back to the user's machine.

b) Building an end-user application to allow literary scholars to carry out text-mining. An application called *noravis* has been built that allows a scholar to mark documents from a collection based on some criteria that the scholar believes important. These labels are then used as input for a text-mining algorithm that carries out a two-class classification. For example, someone studying Emily Dickinson's letters and poems might label a set of documents as being erotic or not-erotic. The *noravis* application invokes text-mining on the server systems, and then presents the scholar with a complete classification of all documents as erotic or not-erotic, along with results showing which vocabulary items were strong markers of eroticism. The *noravis* application is designed to support iteration and exploration by the scholar.

c) Exploring the usefulness of text-mining in support of literary research. We have worked with literary scholars on several research questions that concern them. In addition to the study of eroticism in Dickinson's works described above, other scholars have studied a characteristic known as sentimentalism in 19th century novels.

d) User interfaces for carrying out text-mining tasks and exploring text-mining results. Work in this area has included observation and evaluation sessions with scholars using the *noravis* application, as well as creating initial prototypes of future software tools to be developed in this domain.

In this presentation, I will give an overview of the *nora* project, its products and its activities. I will discuss various lessons-learned related to the processing of our texts and also software architecture. I will give examples of the kinds of humanities-focused research questions that we feel can be addressed by text-mining, and describe some these results that have excited humanities scholars associated with the project.

Teaching a computer to read Shakespeare: the problem of spelling variation

Dawn Archer (UCLAN) and Paul Rayson (Lancaster University)

A number of text mining tools exist, but their application becomes problematic when applied to historical data. There are many reasons as to why this is the case, but the most troublesome is probably that of spelling variation. In this presentation, we explain how we have sought to overcome the spelling issue as part of our work in respect to the development of an historical semantic tagger. The historical tagger is based on an existing semantic tagger, the UCREL Semantic Annotation System, which automatically tags modern English data (spoken and written) with semantic information. More specifically, we will explain how we have developed a VARIant Detector (henceforth VARD) as a means of detecting and normalising spelling variants to their modern equivalent so that USAS can begin to annotate historical data from Shakespeare onwards. We will also demonstrate a new version of VARD (with a new graphical interface developed by Alistair Baron, Lancaster University).

The advantages of using relational databases for large historical corpora

Mark Davies (Brigham Young University)

In the past five years, I have created three large historical corpora that are available on the web: the 100 million word Corpus del Español (<http://www.corpusdelespanol.org>), the 45 million word Corpus do Português (<http://www.corpusdoportugues.org>), and the 37 million word Corpus of Historical English (<http://view.byu.edu/che>). This is in addition to my redesigned architecture and interface for the British National Corpus (<http://view.byu.edu>). All of these corpora use a relational database architecture that allows for size (hundreds of millions of words) and speed (typically less than one or two seconds for most query). More importantly, it allows for a wide range of queries, including word and phrase, substrings, part of speech, lemma, synonyms, collocates, comparison of collocates of related words, frequency in historical periods and genres, and user-defined lists, among others. In addition, because the relational databases include tables with frequency and distributional information for all words in the corpus (and which can easily be queried and updated), it is ideal for efficiently tagging and lemmatizing historical corpora, in cases where there is a great deal of spelling variation.

Digitisation of historical texts at ProQuest and ways of accessing variant word forms

Tristan Wilson (ProQuest Information and Learning)

ProQuest Information and Learning produce a number of web-based historical full-text databases, among them Early English Books Online and Literature Online. Some data is manufactured by us, some by partners. Ours is produced by manual keying or Optical Character Recognition, and stored in a hierarchical format, usually SGML.

The presentation will focus on variant word-forms, which can be a significant hindrance to comprehensive searches. The existing tools for dealing with this, browse or 'look for' lists, are valuable but have their shortcomings. We can handle general typographical variants by replacing likely letters automatically in search terms, but that's a bit clumsy and solves only a limited part of the problem. One promising approach is use of variant spelling databases. We want to keep up with present work on text mining, both to see what techniques or data have proven effective and to find out what our customers would like modern databases to offer.

Lessons Learned from Transcribing and Tagging the Newcastle Electronic Corpus of Tyneside English (NECTE)⁶.

Joan Beal (University of Sheffield) and Nicholas Smith (Lancaster University)

The NECTE corpus presented a number of problems not encountered by those producing corpora of standard varieties. The primary material consisted of audio recordings which needed to be orthographically transcribed and grammatically tagged. Preston (1985), Macaulay (1991), Kirk (1997) and Beal (2005) all note that representing vernacular Englishes orthographically, e.g. by using 'eye dialect' can be problematic on various levels. Apart from unwelcome associations with negative racial or social connotations, there are theoretical objections to devising non-standard spellings which represent certain groups of vernacular speakers, thus making their speech appear more differentiated from mainstream colloquial varieties than is warranted. In the first half of this paper, we outline the principles and methods adopted in devising an Orthographic Transcription Protocol for a vernacular corpus, and the challenges faced by the NECTE team in practice. Protocols for grammatical tagging have likewise been devised with standard varieties in mind. In the second half, we relate how existing tagging software (Garside and Smith 1997) was adapted to take account of the non-standard grammar of Tyneside English.

References

- Beal, J. C. (2005) 'Dialect Representation in Texts' in *The Encyclopedia of Language and Linguistics* 2nd . edn. Elsevier, 531-8.
- Garside, R. and N. Smith. (1997) A hybrid grammatical tagger: CLAWS4. In Garside, R., G. Leech, A. McEnery (eds.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 102-121. London: Longman.
- Kirk, J. (1997) 'Irish-English and contemporary literary writing', in Kallen, J. (ed.) *Focus on Ireland*, 190-205. Amsterdam: John Benjamins.
- Macaulay, R. K. S. (1991) "'Coz it izny spelt when they say it": displaying dialect in writing', *American Speech*, 66: 280-291.
- Preston, D. (1985) 'The Li'l Abner syndrome: written representations of speech', *American Speech*, 60 (4): 328-336.

⁶ www.ncl.ac.uk/necte

Nineteenth Century Serials Edition Project

Suzanne Paylor, Jim Mussell (Birkbeck College)

The move towards digitisation, both for the purposes of preservation and in order to develop new possibilities for historical research, is producing vast digital repositories of nineteenth-century printed matter. Typically consisting of searchable uncorrected OCR and facsimile images of the scanned pages these offer users the opportunity to interact with this material in new and exciting ways. However, if we are to exploit these fully scholars must develop tools for analysis and exploration which also help users to comprehend them as historical objects. We briefly explore the ways in which text mining might address this need by illuminating relationships between form and content, and outline the particular challenges faced by those applying such techniques to nineteenth-century serial literature.

The LICHEN Framework: A new toolbox for the exploitation of corpora

Lisa Lena Opas-Hänninen, Tapio Seppänen, Ilkka Juuso and Matti Hosio
(University of Oulu, Finland)

The Linguistic and Cultural Heritage Electronic Network (LICHEN) project focuses on the languages and cultures of the northern circumpolar region. The project aims to collect, preserve and disseminate information about the languages spoken there, thus also enabling research on them. Secondly, we are creating an electronic framework for the collection, management, online display, and exploitation of existing corpora of these languages, which is also applicable to other corpora that represent other varieties of languages.

Humanities scholars have studied linguistic, educational and social questions related to minority speakers but have been held back by the inability to process and analyse large quantities of data in an effective manner. Although a number of tools have been developed, they suffer from various restrictions, e.g. applicability is restricted, importing data is laborious, user interfaces and encoding standards are outdated, no support for multilinguality is included, or they promise more than they offer.

The LICHEN framework will address all these problems. It is intended to be the equivalent of an extendable toolbox for corpus linguists. The framework will attempt to offer much-needed functionality in an easy-to-use package, which is shaped and built-on according to real user needs. Initially emphasis will be given to the implementation of the text capabilities of the system, but other modalities (such as audio and video) will follow. The idea is to facilitate queries into a multimodal database (i.e. one that can handle text, sound, pictures, video, etc.) using both proven and novel ways of finding and displaying information. Metadata and metadata visualisation, particularly in conjunction with the new modalities, will be essential in achieving this. Tools and experiences from work carried out at the MediaTeam Oulu research group (Faculty of Engineering, University of Oulu) will be utilized here.

The framework will make migration of data to and from other tools straightforward by offering import and export features for commonly used programs. It will enable users to bring in their own data, which they can keep private or make public using the built-in web functionality, if they so wish. The database will also be capable of handling several different versions of any document (for example, revisions, interpretations or translations); these are linked, a feature that can be made use of in queries. Queries can be made using regular expressions, which may combine free-form text (words, phrases) and part-of-speech tags, for example.

The framework will be implemented in the Java programming language making it platform independent and taking advantage of the many technology components developed for that language. Support for multiple languages and a variety of character encoding schemes will be important. New features can be added to any installation of the framework by downloading the desired toolboxes via a web update feature. We aim to establish and follow best engineering practices in all aspects of the framework.

We welcome the input of the linguistic research community into shaping the tools we are developing since our aim is to create something truly useful.

Participant list

Jean Anderson	University of Glasgow
Dawn Archer	University of Central Lancashire
Alistair Baron	Lancaster University
Joan Beal	University of Sheffield
Carmela Chateau	University of Burgundy
Ruzanna Chitchyan	Lancaster University
Jonathan Culpeper	Lancaster University
Mark Davies	Brigham Young University
Astrid Ensslin	University of Manchester
Andrea Ernst-Gerlach	Universität Duisburg-Essen
Ricardo Gacitua	Lancaster University
Maria F. Garcma-Bermejo Giner	University of Salamanca
Sharon Givon	Edinburgh University
Barry Haddow	University of Edinburgh
Andrew Hardie	Lancaster University
Marian Hellema	National Library of the Netherlands
Markus Heller	Munich University
Tom Horton	University of Virginia
Dolores Iorizzo	Imperial College Lonon
Meng Ji	Imperial College Lonon
Samuli Kaislaniemi	University of Helsinki
Christian Kay	University of Glasgow
Maggie Kerridge	University of Sussex
Merja Kytö	Uppsala University
Geoffrey Leech	Lancaster University
Jim Mussell	Birkbeck College London
Lisa Lena Opas-Hänninen	University of Oulu
Suzanne Paylor	Birkbeck College London
Yves Peirsman	University of Leuven
Wim Peters	University of Sheffield
Thomas Pitz	Universität Duisburg-Essen
Sheryl Prentice	Lancaster University
Stephen Pumfrey	Lancaster University
Paul Rayson	Lancaster University
Brian Rea	University of Liverpool
Rob Sanderson	University of Liverpool
Mike Scott	University of Liverpool
Nick Smith	Lancaster University
Dan Smith	University of Manchester
Reiko Takeda	King's College London
Tristan Wilson	Proquest UK
May Lai-Yin Wong	University of Hong Kong