

Colloquium to be held at ISFC 2008 Sydney 2008

<http://minerva.ling.mq.edu.au/isfc/>

Title:

SFL Corpus and Computing Colloquium

Convenors:

Sabine Bartsch und Elke Teich, Technische Universität Darmstadt, Department of Linguistics and Literary Studies, English Linguistics, Darmstadt, Germany, e-mail: {lastname}@linglit.tu-darmstadt.de, URL: <http://www.linglit.tu-darmstadt.de>

Maria Herke, Macquarie University, Department of Linguistics, Sydney, Australia, e-mail: maria.herke@ling.mq.edu.au, http://www.ling.mq.edu.au/about/staff/herke__maria/index.html

Mick O'Donnell, Universidad Autónoma de Madrid, Informatics Department, Madrid, Spain, e-mail: micko@wagsoft.com, URL: <http://arantxa.ii.uam.es/~modonnel/>, <http://www.wagsoft.com>

Canzhong Wu, Macquarie University, Department of Linguistics, Sydney, Australia, e-mail: canzhong.wu@ling.mq.edu.au, URL: http://www.ling.mq.edu.au/about/staff/wu_canzhong/index.html

Timetable for workshops and colloquia, the time table of activities or presentations within the 3:30 hour slot:

Time	Duration	Activities / presentations	Presenter
00:00	45	C1 Opening talk: SFL methodology	Elke Teich
00:45	30	C2 Corpus-based SFL: corpora and annotations	Sabine Bartsch
01:15	15	SHORT BREAK	
01:30	30	C3 Tools for sustainable SFL language resources	Mick O'Donnell
02:00	30	C4 Tools for multilingual / translation research in SFL	Canzhong Wu
02:30	30	C5 Corpus-based SFL research	Maria Herke
03:30	30	C6 Wrap-up: final discussion	panel (all)

Presentational requirements:

data projector, (wireless) online access, whiteboard

Presentation: Aim Succinct statement of the aim of the contribution

The aim of this colloquium is to support and continue recent discussion within the SFL community concerning questions of the intersection of SFL theory and modern flavours of empirical methodology as are embodied by corpus-based, computer-aided empirical methodology.

Presentation: Background Relationship to previous and current work within systemic functional linguistics, with key references

The colloquium is to be seen in the broader context of discussions within the SFL community pertaining to the inherently empirical orientation of SFL. These issues appear to be especially important in the light of developments in corpus and computational linguistics in the past 20 years.

This colloquium is furthermore situated in the context of the Initiative for A Repository of SFL Resources (IRSFL) organised by the Halliday Centre, City University of Hong Kong, the Department of Linguistics at Macquarie University, Sydney, Australia and the Department of Linguistics and Literary Studies at Technische Universität Darmstadt, Germany. This initiative is subdivided into three intersecting initiatives which are going to be addressed in this workshop:

- SFL text archive (Christian Matthiessen, Macquarie, Sydney),
- SFL text annotation/corpora (Maria Herke, Macquarie University, Sydney), and
- SFL tool infrastructure (Elke Teich, TU Darmstadt)

In the light of a broader re-orientation of linguistics towards empirical approaches, recent discussions within the SFL community have begun to evaluate the possibility of devising modern corpus and computational methods for SFL as well as developing the respective SFL resources such as text archives, SFL annotated corpora and tools for SFL data processing. The development of such resources, due to the requirements posed by different researcher interests and the effort involved is a community task which requires agreement on the nature of the resources in terms of coverage, diversity and standards enabling exchange and comparability of data.

Presentation: Body Description of the main part of the presentation

From its inception, SFL has been empirical in orientation. This orientation becomes evident in the wealth of individual studies of linguistic data (see e.g. Hasan's (1986) work on mother-child language, Martin's research, as well as studies of individual languages other than English such as Caffarel on French and Teruya on Japanese). Yet, while the value of these in-depth qualitative studies and descriptions based on SFL theory cannot be estimated high enough, one drawback of a methodological approach focused on highly individual studies of data is that this data is often private or proprietary in nature, but at any rate not made available to a wider community. The data does thus not become accessible for inspection or analysis by other members of the research community, but is only indirectly presented through paper or book publications.

The aim of this colloquium is to help promote discussion and commitment within the SFL community concerning the compilation of text archives covering the multitude of languages beyond English to which SFL is applied especially in recent years as well as the establishment of standardised formats of corpus compilation and annotation which help ensure the exchange of annotated resources. A central prerequisite for these efforts is addressed and promoted by the third area of concern of this colloquium, the development and establishment of tools suitable for the work of a wider SFL community. To this end, the colloquium is comprised of talks presenting current trends in corpus-based SFL research and reporting on experience from corpus-based research within an SFL framework. To complement this program, the colloquium invites contributions to the discussion within the colloquium from SF linguists who are either working within a corpus-driven or computational linguistic framework, have been involved in the collection of text archives or have experience in the systematic compilation and annotation of SFL resources.

Presentation: References List of references cited in the abstract

Caffarel, Alice. (2006). *A systemic functional interpretation of French grammar: from grammar to discourse*. Continuum Publishing: London.

Halliday, M.A.K. (1996). On Grammar and Grammaticals. In R. Hasan, C. Cloran, D. Butt (eds.), *Functional Descriptions Theory in Practice*. Amsterdam: John Benjamins. p. 24 ff.

Hasan, Ruqaiya. (1986). 'The ontogenesis of ideology: an interpretation of mother child talk'. In: Terry Threadgold, E.A. Grosz, Gunther Kress and M.A.K. Halliday (eds.) *Language, Semiotics, Ideology*. (Sydney Studies in Society and Culture 3). The Sydney Association for Studies in Society and Culture. pp.125-146.

Teruya, Kazuhiro. (2006). *A Systemic Functional Grammar of Japanese*. Continuum Publishing: London.

Thompson, Geoff, Susan Hunston. (eds.) (2006). *System and Corpus: Exploring Connections*. London: Equinox.

Presentation: Key terms Key terms identifying the field(s) of presentation

SFL, corpus linguistics, corpora, annotation, tools

Abstracts from the convenors of the proposed colloquium

C1 Opening talk: SFL methodology	Elke Teich
C2 Corpus-based SFL: corpora and annotations	Sabine Bartsch
C3 Tools for sustainable SFL language resources	Mick O'Donnell
C4 Tools for multilingual / translation research in SFL	Canzhong Wu
C5 Corpus-based SFL research	Maria Herke

C1 SFL Methodology

From (no) hypothesis to analysis and interpretation: a data-based, theory-driven methodology

Elke Teich

The paper describes possible corpus-based analysis designs in the context of Systemic Functional Linguistics. I discuss two starting positions: (a) there is an initial hypothesis, and (b) there is no particular initial hypothesis. Depending on the starting position, there are different methods of data exploration/analysis, which will be introduced in the paper. Finally, the necessary computational tools to carry out SFL-driven corpus analyses will be discussed.

C2 Corpus-based SFL: corpora and annotations

Corpus-based SFL: A proposal for corpora and annotations

Sabine Bartsch

Despite being inherently empirical in orientation, SFL has not yet fully begun to explore the possibilities offered through recent advances in corpus and computational linguistics, especially corpora annotated to cater for specific research questions. This paper investigates current corpus and annotation designs and attempts to show which types of corpus design and annotation may be suitable for the research questions addressed by SFL.

The paper takes as its vantage point a subset of research questions typically addressed by systemic functional linguistics and explores the desiderates to the construction and annotation of corpora suitable in order to explore these questions by means of conventional and computational approaches.

C3 Tools for sustainable SFL language resources

Automating SFL Annotation: what we can do, what is left to do

Mick O'Donnell

SFL currently lacks reliable parsers for systemic grammars, and other areas of SFL are even more out of reach (appraisal analysis, aspects of discourse analysis). However, some aspects of SFL annotation can be automated. Given the degree of error in this annotation, such automatic annotation processes need to be embedded within a user interface which allows a human to correct the annotation. This paper will explore some of the efforts at automatic annotation of text from an SFL perspective, and how it can be integrated into manual annotation tools to allow correction.

C4 Tools for multilingual research in SFL

Canzhong Wu

Multilingual studies has started to emerge as a new research field for investigating related but traditionally mutually insulated areas such as translation, language typology, second language learning and bilingualism. This paper examines the various tools available in these areas, and explores how they contribute to this emerging field in the context of systemic functional linguistics.

C5 Corpus-based SFL research

Maria Herke