Spoken Language Across Time

INTERNATIONAL SYMPOSIUM | SEPTEMBER 2019 | LUND UNIVERSITY

Main Auditorium, Centre for Languages and Literature, Lund University, 20 September, 2019

The purpose of this symposium is to celebrate the launch of the brand new corpus of spoken English, the London–Lund Corpus 2 (LLC–2) half a century after the launch of the first London–Lund Corpus (LLC–1). To this end, we are pleased to present an excellent line-up of prominent researchers with an interest in the nature of spontaneous speech in real communicative situations.

Arranged with the support of The Birgit Rausing Language Programme and the Royal Society of Letters at Lund

For more information, please visit:
www.sol.lu.se/en/subjects/engelska/research/llc2/

Programme | 09.00-12.00

09.00–09.20 | Welcome
Johannes Persson, Dean of the Joint Faculties of Humanities and Theology

09.20–10.00 | Nele Põldvere, Victoria Johansson and Carita Paradis, Lund University
The new London–Lund Corpus (LLC–2): design, compilation, access

10.00–10.30 | Bas Aarts, University College London
Research in spoken English: the London–Lund experience

10.30–11.00 | Fika

11.00–11.30 | Gunnel Tottie, University of Zurich
Corpus linguistics now and then: the case of negation of indefinites

11.30–12.00 | Karin Aijmer, University of Gothenburg
'They're like proper crazy like' – New uses of intensifiers in spoken English

12.00–13.30 | Lunch break

Programme | 13.30-16.45

13.30–14.00 | Robbie Love, University of Leeds
Building and analysing a national corpus of informal spoken English: the Spoken BNC2014

14.00–14.30 | Susan Reichelt, University of Greifswald
Combining apparent and real time approaches to language change: Recent developments of *kind of* and *sort of* in spoken British English using *BNClab*

14.30–15.00 | Jonathan Culpeper, Lancaster University
On 'spokenness': From Early Modern to Present-day English

15.00–15.30 | Fika

15.30–16.30 | Herbert Clark, Stanford University (Keynote speaker)
On the use and misuse of language corpora

16.30–16.45 | Closing
Professor Emeritus Jan Svartvik

Nele Põldvere, Victoria Johansson and Carita Paradis, Lund University

THE NEW LONDON–LUND CORPUS (LLC–2): DESIGN, COMPILATION, ACCESS

This talk reports on the compilation of the new London–Lund Corpus (LLC–2) – a corpus of contemporary spoken British English, collected 2014–2019. The size and design of LLC–2 are the same as that of the world's first corpus of spoken language, namely the London–Lund Corpus (LLC–1), with spoken data mainly

Nele Põldvere   Victoria Johansson   Carita Paradis

from the 1960s. In addition to the fact that we have a corpus of contemporary speech, the existence of LLC–2 also gives researchers the opportunity to make principles diachronic comparisons of speech over the past 50 years and detect change in communicative behaviour among speakers.

The compilation of LLC–2 has included a number of different stages such as data collection, transcription of the recordings, markup and annotation, and finally making the corpus accessible to the research community. The talk describes and critically examines the methodological decisions made in each stage. For example, it was important to strike a balance between LLC–2 as a representative collection of data of contemporary spoken English and its

comparability to LLC–1. Therefore, both corpora contain the same speech situations (dialogue, mainly everyday face-to-face conversation, as well as monologue), but the specific recordings added to LLC–2 also reflect the technological advances of the last few decades, particularly with respect to speech situations such as telephone calls (e.g., Skype) and broadcast discussions and interviews (e.g., podcasts). Moreover, the transcriptions in LLC–2 are orthographic and time-aligned with the corresponding sound files, which is a feature of the corpus that is novel and makes it possible to, among other things, investigate prosody and dialogue management among speakers with great precision. The corpus, as well as metadata about the transcriptions and the speakers, will be released to the public in late 2019 from the Lund University Humanities Lab's corpus server. The release will fill an unfortunate gap in the availability of spoken corpora for linguistic analysis.

The benefits of spoken corpora in general and of LLC–2 in particular will be demonstrated in the talk through examples of case studies based on the corpus (e.g., Põldvere & Paradis, 2019a, 2019b). The case studies illustrate how LLC–2 can contribute to our understanding of meaning-making and discursive practices in real communication and provide a window into the cognitive and social processes of dialogic interaction, both from a contemporary and a back-in-time perspective.

Bas Aarts, University College London

RESEARCH IN SPOKEN ENGLISH: THE LONDON–LUND EXPERIENCE

In this paper I will go back to the beginning by tracing the history of the collaboration between the Survey of English Usage (which celebrates its 60th anniversary this year) and the University of Lund. I will briefly present the corpus exploration tools that we developed, and how they can be used to carry out research on both written and spoken English. I will then present the results of some recent research in the Survey of English Usage on spoken English, specifically work on the progressive construction, modal verbs and the perfect construction.

Bas Aarts

Gunnel Tottie, University of Zurich

CORPUS LINGUISTICS NOW AND THEN: THE CASE OF NEGATION OF INDEFINITES

I will illustrate the progress (and the woes) of corpus linguistics with examples from my own work on two problems in the syntax of negation in English.

Gunnel Tottie

I first studied the variation between NO-negation and NOT-negation, as in (1) and (2), begining with the Brown Corpus in the 1970s and adding the London–Lund Corpus in the 1980s (Tottie 1983, 1991).

(1) NO-negation: I have **no** dog/money/friends.
(2) NOT-negation with **a** or **any**: I don't have **a** dog/**any** problem/money/friends.

The second problem was impossible to address because of the paucity of material in the seventies and eighties: the variation between the indefinite article *a/an* and *any* as indefinite determiners of count nouns in sentences with NOT-negation, as in (3):

(3) It isn't **a/any** problem/There isn't **a/any** problem/I don't have/see **a/any** problem.

With the advent of mega-corpora like the *Corpus of Contemporary American English* (COCA), comprising 577 million words, it was tempting to try to study the variation between *a/n* - or *any*-negation and try to find the factors conditioning their use. This is what I am currently working on – not without complications, due both to the sheer size and the makeup of the corpus.

Karin Aijmer, University of Gothenburg

'THEY'RE LIKE PROPER CRAZY LIKE' – NEW USES OF INTENSIFIERS IN SPOKEN ENGLISH

New intensifiers emerge, become fashionable but can then lose their popularity and be replaced by other more striking intensifiers. When Paradis (2000) revisited degree modifiers of adjectives in the 1990's, she found, for example, that the intensifiers in the London–Lund Corpus occurred with different frequency in the COLT corpus (the Bergen Corpus of London Teenagers) that she used for comparison. Moreover, she found some examples of 'new' intensifiers such as *well* (*well weird*) and *enough* (*enough bad*). The aim of my presentation is to discuss changes in the intensification system which have taken place after this. The focus will be on some intensifiers (eg *well, all, proper, pretty*) which seem to have become more frequent or have emerged recently in spoken British English. The material for this study is taken from the spoken British National Corpus 2014 (Love et al 2017). On-going changes can be observed by comparing the distribution and uses of the same intensifiers in the old BNC (BNC1994). The research questions are:

– What are the mechanisms responsible for the changes?
– What is the role of sociolinguistic factors such as the age and gender of the speakers to explain the changes?

Karin Aijmer

Robbie Love, University of Leeds

BUILDING AND ANALYSING A NATIONAL CORPUS OF INFORMAL SPOKEN ENGLISH: THE SPOKEN BNC2014

The Spoken BNC2014 (Love et al. 2017, Love forth.) is an important component of the new British National Corpus 2014;[i] a large dataset representing current British English usage across different situations, which is being compiled by Lancaster University in collaboration with Cambridge University Press. It is the successor to the spoken component of the original British National Corpus (Crowdy 1995) and was released publicly via Lancaster University's CQPweb server (Hardie 2012) in September 2017.

In terms of corpus construction, I pay attention to other contemporary spoken corpus projects such as the London–Lund Corpus 2 (Paradis et al. 2015–), the spoken component of CorCenCC (Knight et al. 2016) and FOLK (Schmidt 2016) and consider the role of representativeness in corpus design. I argue that representativeness is an ideal but that it is inevitable – due to practical constraints – that there will be some differences between the original design of a large 'national' corpus and the finished product, and that it is important to be honest, critical and realistic about representativeness. I then demonstrate the research potential of the Spoken BNC2014 with examples from recent research into adverbs (Goodman & Love 2019).

Robbie Love

[i] http://cass.lancs.ac.uk/bnc2014/

Susan Reichelt, University of Greifswald

COMBINING APPARENT AND REAL TIME APPROACHES TO LANGUAGE CHANGE: RECENT DEVELOPMENTS OF *KIND OF* AND *SORT OF* IN SPOKEN BRITISH ENGLISH USING *BNCLAB*

This study reports on ongoing changes in the use of hedges *sort of* and *kind of* in spoken British English of the past twenty years. Following known sociolinguistic patterns of change in progress (c.f. Bailey, 2008; Pichler et al., 2018), special focus will be put on three categories of time: age, date of birth, and date of corpus compilation.

Susan Reichelt

The data used in this study stem from two subsets of the original BNC from 1994 and the newly compiled BNC2014. Both sets were, where possible, balanced across social categories of age, gender, location, and social class. Feature tokens were extracted using the online platform *BNClab*, which includes a concordance viewer alongside first data evaluations, visualizations, and teaching materials. The design of the subsets, highlighted further in this presentation, allows for a combination approach to change, using apparent time and real time trend analyses.

The features under investigation, hedges *sort of* and *kind of*, are often treated as having "basically the same meaning" (Mauranen, 2004: 179; see also Aijmer, 1984: 118), yet show distributional differences across different varieties of English. In British English context, *sort of* is often found as more dominant (cf. Aijmer, 1984; Biber et al., 1999; Gries & David, 2007; Kay, 1984; Mauranen, 2004). Feature use within the two BNC datasets suggests that the variants are currently undergoing change. *Kind of* is increasingly encroaching on *sort of* – a change that becomes observable through the inclusion of the three categories of time, as mentioned above.

The talk thus highlights the need and usefulness of corpora that allow the researcher to combine apparent and real time approaches in order to gain a full picture of ongoing linguistic change.

Jonathan Culpeper, Lancaster University

ON 'SPOKENNESS': FROM EARLY MODERN TO PRESENT-DAY ENGLISH

This paper reflects on 'spokenness' in English from the early modern period to today. I begin by (a) making some general remarks on spokenness in the history of English, and (b) introducing a descriptive approach to 'writtenness' and 'spokenness', one revolving around three categories, namely, the degree to which a text is speech-like, speech-based or speech-purposed. This approach was part of the corpus-based work on spoken interaction in historical English writing that I conducted over 20 years with Merja Kytö (e.g. Culpeper and Kytö 2010). I discuss some of the problems we encountered and some of our findings, in particular those relating to what we termed 'pragmatic noise' (essentially, primary interjections, the noises – ooh's and aah's – that carry pragmatic meanings). I identify the five pragmatic noise items that occurred most frequently in all our speech-related genres but hardly occurred in our non-speech-related genres, and also briefly account for their development. In addition, I discuss at some length the case of the genre of play-texts, a complex hybrid spoken-written genre. Using corpus-based methods, I show how it has changed over the centuries, and relate some of those changes to changes in context.

Jonathan Culpeper

Herbert Clark, Stanford University

ON THE USE AND MISUSE OF LANGUAGE CORPORA

A corpus of English conversation, edited by Jan Svartvik and Randolph Quirk, was a pioneering work when it appeared in 1980. Although many corpora have been published since then, the London–Lund Corpus (LLC) remains unique in its quality and detail. It has proven especially valuable for studying features of the rough and tumble language of spontaneous conversation. Still, like most language corpora, the LLC had to ignore the gestural side of language use, even though gestures are an essential part of deictic expressions such as *I, you, we, those guys, over there, the other side, now*, and *a moment ago*. The problem is that many investigators have proceeded as if language use were all speech and no gesture, and that has left their models wanting.

Herbert Clark

· REGIA · ACADEMIA · CAROLINA ·

www.sol.lu.se/en/subjects/engelska/research/llc2/

LUND UNIVERSITY

SOL
Helgonabacken 12
SE-223 62 Lund
www.sol.lu.se