



INTERNATIONAL SOCIETY for the LINGUISTICS of ENGLISH

SPONSORED BY
THE INTERNATIONAL SOCIETY FOR THE LINGUISTICS OF ENGLISH
THE NEW ENGLAND COMMITTEE FOR ISLE 2
THE ENGLISH DEPARTMENT AT BOSTON UNIVERSITY
THE BOSTON UNIVERSITY HUMANITIES FOUNDATION

plenary speakers

LISA GREEN

UNIVERSITY OF MASSACHUSETTS

Multiple Grammars and Dialectal Variation: A View from the Perspective of Language Development

STEFAN GRIES

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

The Quantitative Revolution in Corpus Linguistics: Applications and their Theoretical Implications

APRIL MCMAHON

THE UNIVERSITY OF EDINBURGH

Comparing [ɹaɪk] with [ɹaɪk]: Methods for Collecting and Comparing Data from Varieties of English

CHRISTIAN MAIR

ROBERT LUDWIGS-UNIVERSITÄT FREIBURG

World Non-Standard Englishes: Reflections on the Global Spread of (Some) Vernacular Varieties of English

CHRISTOPHER RICKS

BOSTON UNIVERSITY

The Very Words, and Not Only Those

SALI TAGLIAMONTE

UNIVERSITY OF TORONTO

System and Society in the Evolution of Change: The Case of Canada

presidential address

DAVID DENISON

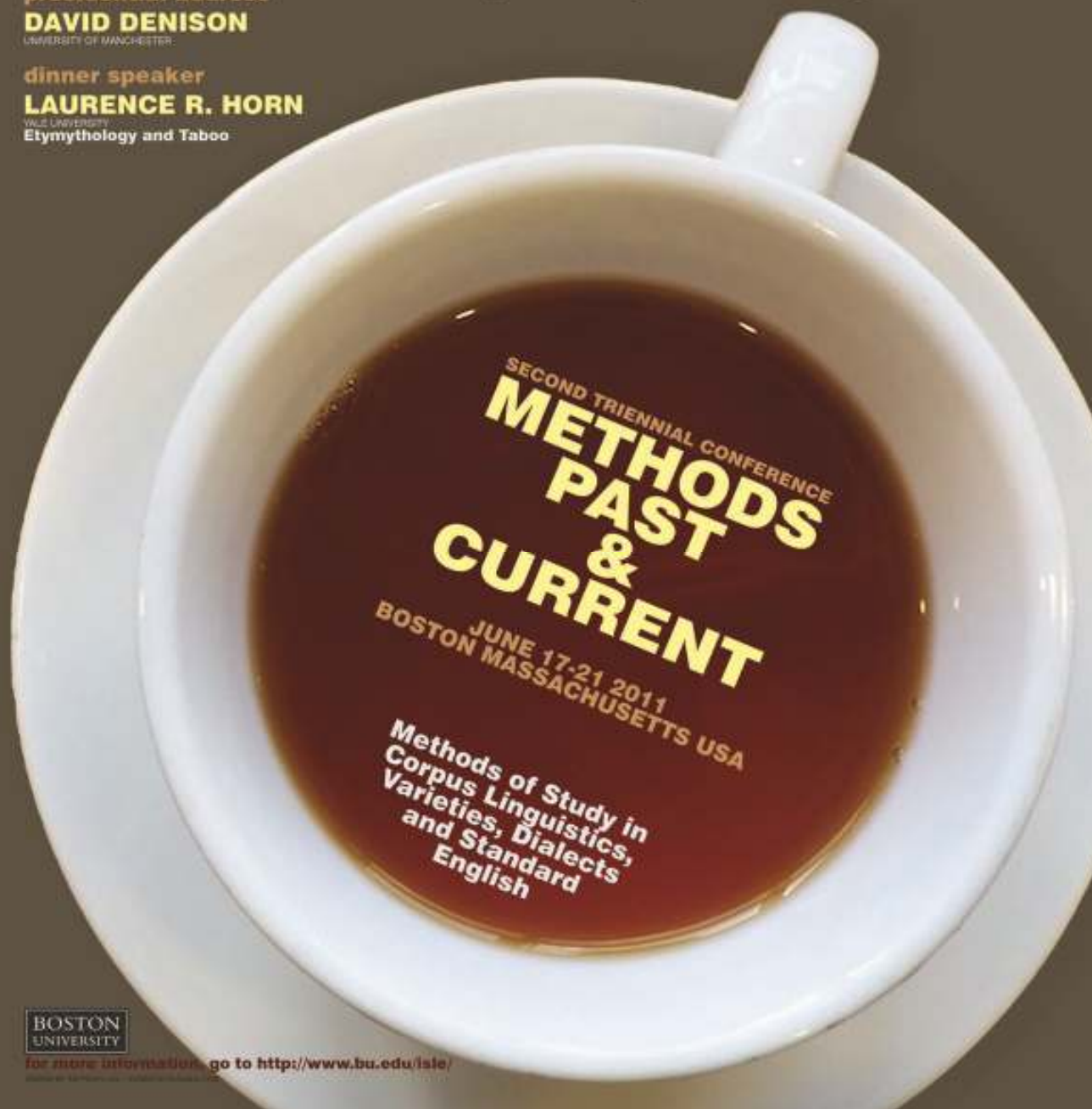
UNIVERSITY OF MANCHESTER

dinner speaker

LAURENCE R. HORN

YALE UNIVERSITY

Etymology and Taboo



for more information, go to <http://www.bu.edu/isle/>

Contents

ISLE2 Conference Schedule	3-13
Schedule Table	15-20
Plenary Abstracts	21-27
Abstracts for Talks	28-91
Poster Abstracts	92-98
Workshop Abstracts	99-146
Transportation	147
Campus Facilities (Internet, Fitness)	147
Restaurants Near BU	148-149
Boston Activities	150-151
Contact Information	152
BU Campus Map	153
Index of Speakers/Authors	154-157

ISLE 2011 Conference Schedule

Unless otherwise noted below, all events are held at Boston University's College of Arts and Sciences (CAS), located at 725 Commonwealth Avenue, Boston.

Please visit our exhibitors' room, located in CAS 216, during conference hours Friday-Monday.

ISLE volunteers are available to assist you in CAS 220 during conference hours.

THURSDAY, JUNE 16

4:00-8:00 REGISTRATION (CAS 116)

8:00-10:00 RECEPTION AT BOSTON UNIVERSITY PUB (225 Bay State Road, behind CAS and in the basement of "The Castle")

FRIDAY, JUNE 17

8:00-9:00 REGISTRATION (CAS 116)

9:00-9:15 OPENING REMARKS; GREETING FROM DEAN PATRICIA JOHNSON (CAS 224)

9:15-10:15 PLENARY SESSION (CAS 224)
Christian Mair, "World Non-Standard Englishes: Reflections on the Global Spread of (Some) Vernacular Varieties of English" (Chair: Charles Meyer)

10:15-10:45 COFFEE BREAK (CAS 2nd floor hall)

General Session I

Section A (CAS 203)

Internet idioms (Chair: Daniel Donoghue)

10:50-11:20 Jon Bakos, "QQ More"

11:25-11:55 Daphné Kerremans and Susanne Stegmayr, "Neologisms on the internet"

12:00-12:30 Ursula Kirsten, "Development of SMS language from 2000 to 2010"

12:30-1:55 LUNCH

Case (Chair: Bas Aarts)

2:00-2:30 John Payne and Eva Berlage, "The effect of semantic relations on genitive variation"

2:35-3:05 Christoph Wolk, Joan Bresnan, Anette Rosenbach and Benedikt Szmrecsanyi, "Dative and genitive variability in late ModE"

3:10-3:40 Stefanie Wulff and Stefan Th. Gries, "A multifactorial study of genitive alternation in L2 English"

3:45-4:10 COFFEE BREAK (CAS 2nd floor hall)

Perceptual Dialectology (Chair: Bas Aarts)

4:15-4:45 Chris Montgomery, "A new method for dialect recognition and rating in perceptual dialectology"

Section B (CAS 237)

Comparative studies of Modern British and American Constructions (Chair: Dagmar Deuber)

10:50-11:20 Thomas Hoffmann: "The more Data, the better"

11:25-11:55 Gunther Kaltenböck, "Comment clauses on the move"

12:00-12:30 Turo Vartiainen, "Conceptual proximity and the positional variation of directional modifiers in English"

12:30-1:55 LUNCH

African and related diasporic Englishes (Chair: Gunther Kaltenböck)

2:00-2:30 Lars Hinrichs, "Gauging variety status in diasporic dialect mixing"

2:30-3:05 Magnus Huber and Sebastian Schmidt, "New ways of analysing the history of varieties of English. Early Highlife recordings from Ghana"

3:10-3:40 Robert Fuchs, "The progressive aspect in Nigerian English"

3:45-4:10 COFFEE BREAK (CAS 2nd floor hall)

4:15-4:45 Glenda-Alicia Leung, "Approaching the Acrolect"

Section C (CAS 316)

Irish English (Chair: Lauren Hall-Lew)

10:50-11:20 Julia Davydova, "Detecting historical continuity in modern Singapore English: A case study of the present perfect"

11:25-11:55 Marije van Hattum, "A preparation of news to come in Irish immigrant letters"

12:00-12:30 Stephen Lucek, "Invariant tags in Irish English"

12:30-1:55 LUNCH

Phonological Topics in American English and New Englishes (Chair: Katie Drager)

2:00-2:30 David Eddington, "Flaps and other variants of /t/ in American English"

2:35-3:05 Caroline Wiltshire, "New Englishes and the emergence of the unmarked"

3:10-3:40 Toshihiro Oda, "Phonetically accidental and systematic gaps"

3:45-4:10 COFFEE BREAK (CAS 2nd floor hall)

Section D (CAS 226)

Variationism (Chair: Lynn Clark)

10:50-11:20 Don Chapman, "Why empirical studies of prescriptive rules should be variationist"

11:25-11:55 Kirk Hazen, "Morphological methodology for a rapidly reconfigured variable"

12:00-12:30 Sandra Jansen, "Variation and Change in the north-west of England"

12:30-1:55 LUNCH

2:00-2:30 Presley Ifukor, "Towards the emergence of technolectal Nigerian English"

Academic Styles (Chair: Lynn Clark)

2:35-3:05 Ute Römer, "The phraseological profile model applied: New insights into academic speech and writing"

3:10-3:40 Peter Siemund, "Varieties of English in the classroom"

3:45-4:10 COFFEE BREAK (CAS 2nd floor hall)

4:55-5:55 PLENARY SESSION (CAS 224)

David Denison, Presidential Address (Chair: Elizabeth Traugott)

SATURDAY, JUNE 18

8:30-9:30 PLENARY SESSION (CAS 224)

April McMahon, "Comparing [laɪk] with [lʌɪk]: Methods for Collecting and Comparing Data from Varieties of English" (Chair: Stephen Harris)

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

General Session II

Section A (CAS 213)

Workshop: Kevin Watson, Lynn Clark, Warren Maguire: Mergers in English:

Perspectives from phonology, sociolinguistics and psycholinguistics

10:10-10:40 Warren Maguire, Lynn Clark, and Kevin Watson, "The meaning of 'merger'"

10:45-11:15 Maciej Baranowski, "On the role of social factors in vocalic mergers"

11:20-11:50 Lynn Clark and Kevin Watson, "Capturing listeners' real-time reactions to the NURSE~SQUARE merger"

11:55-12:25 Katie Drager and Jennifer Hay, "Mergers in production and perception"

12:30-1:55 LUNCH

2:00-2:30 Lauren Hall-Lew, "Interpreting 'flip-flop' patterns in vowel mergers-in-progress"

2:35-3:05 Jennifer Nycz, "New contrast acquisition: Methodological issues & theoretical implications"

3:05-3:35 EXHIBITORS' COFFEE HOUR (CAS 216)

3:40-4:10 Phillip Tipton, "Modelling (socio)linguistic mergers: the role of global context in the processing of social and linguistic information"

Section B (CAS 237)

Modern English constructions (Chair: Stefan Diemer)

10:10-10:40 Bas Aarts, Jill Bowie, and Sean Wallis, "Typical and atypical change in modal usage over time"

10:45-11:15 Karin Axelsson, "A new functional model for tag questions based on fiction dialogue data"

11:20-11:50 Linnea Micciulla, "Factbors predicting the use of passive voice in newspaper headlines"

12:30-1:55 LUNCH

Pragmatics (Chair: Markus Bieswanger)

2:00-2:30 Markus Bieswanger, "Variationist sociolinguistics meets variational pragmatics"

2:35-3:05 Christine Günther, "Pragmatic factors determining variation in the realization of head nouns"

3:05-3:35 EXHIBITORS' COFFEE HOUR (CAS 216)

3:40-4:10 Meike Pfaff, "On the pragmatics of obligative *want to*"

4:15-4:45 Alexander Bergs, "On how to integrate context into grammar"

Section C (CAS 316)

Workshop: John Payne and Eva Berlage: Genitive variation in English

10:10-10:40 John Payne and Eva Berlage, "Genitive variation: the role of the oblique genitive"

10:45-11:15 Sali Tagliamonte and Bridget Jankowski, "On the genitive's trail: data and method from a sociolinguistic perspective"

11:20-11:50 Cathy O'Connor, "Is animacy the most important factor in predicting the English possessive alternation?"

11:55-12:25 Kersti Börjars, David Denison and Grzegorz Krajewski, “Poss-s vs. poss-of revisited”

12:30-1:55 LUNCH

2:00-2:30 Katharina Ehret, Christoph Wolk, and Benedikt Szmrecsanyi, “Genitive variation in Late Modern English: focus on weight and rhythm”

2:30-3:05 Evelien Keizer, “Pre- and postnominal possessives in English, Dutch and German – an FDG account ”

3:05-3:35 EXHIBITORS' COFFEE HOUR (CAS 216)

Canadian English (Chair: Daniel Donoghue)

3:40-4:10 Charles Boberg, “Ethnicity and regional variation in Canadian English”

4:15-4:45 Stefan Dollinger: “New Dialect Formation cum Dynamic Model: Language attitudes and the case of Vancouver English

Section D (CAS 324)

Workshop: Marianne Hundt: English in the Indian Diaspora

10:10-10:40 Dagmar Deuber, Glenda Leung and Véronique Lacoste, “Indo-Trinidadian speech: features and stereotypes”

10:45-11:15 Marianne Hundt, “Zero articles in Indian Englishes: a comparison of primary and secondary diasporasituations”

11:20-11:50 Jakob R. E. Leimgruber, “Singapore’s Indian community: linguistic, social, and sociolinguistic aspects”

11:55-12:25 Rajend Mesthrie, “The making of a dialect dictionary 1: where does a New English dictionary stop?”

12:30-1:55 LUNCH

2:00-2:30 Claudia Rathore, “East African Indians in Leicester, UK: phonological variation across generations”

2:35-3:05 Farhana Alam and Jane Stuart-Smith, “Identity, ethnicity and fine phonetic detail: an acoustic phonetic analysis of syllable-initial /t/ in Glaswegian girls of Pakistani heritage”

3:05-3:35 EXHIBITORS' COFFEE HOUR (CAS 216)

3:40-4:10 Lena Zipp, “Features of IndoFijian English across registers”

4:15-4:45 Capstone Session

4:55-5:55 PLENARY SESSION (CAS 224)

Christopher Ricks, “The very words, and not only those” (Chair: Daniel Donoghue)

SUNDAY, JUNE 19

8:30-9:30 PLENARY SESSION (CAS 224)

Sali Tagliamonte, University of Toronto: "System and society in the evolution of change: The case of Canada" (Chair: Laurel Brinton)

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION III

Section A (CAS 213)

Corpus Studies (Chair: Magnus Huber)

10:10-10:40 Garrison Bickerstaff, "Flexibility and application of the bounded virtual corpus"

10:45-11:15 Terttu Nevalainen, "Tools for comparing corpora"

11:20-11:50 Matthew O'Donnell, "The adjusted frequency list"

Case Studies

11:55-12:25 Lieven Vandelanotte, "*Call so and so and tell him such and such*: A corpus-based study of suspensive reference in contemporary English"

12:30-1:55 LUNCH

Case Studies (Chair: Heli Paulasto)

2:00-2:30 Gregory Garretson, "A new perspective on antonymy"

2:35-3:05 Stefan Diemer, "Corpus linguistics with Google?"

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Michael Erlewine, "The Constituency of Hyperlinks in a Hypertext Corpus"

Section B (CAS 237)

Workshop: Lars Hinrichs and Stefan Dollinger: Aspects of methodology and pedagogy

A. Lars Hinrichs and Stefan Dollinger: Long-term research projects on local varieties of English

10:10-10:40 Walt Wolfram, "The Theoretical and Methodological Challenge of Longitudinal Studies: The Case of African American English"

10:45-11:15 Thomas Purnell, Eric Raimy and Joseph Salmons, "The Wisconsin Englishes Project and WiSCO"

11:20-11:50 Bill Kretzschmar, "Student Participation in the Linguistic Atlas Project"

11:55-12:25 Kirk Hazen, "Goals for the project and your career: Long term success"

12:30-1:55 LUNCH

Workshop: B. Marnie Reed: Evaluation and Instruction

2:00-2:30 Jarek Weckwerth, "Variation in the production of the TRAP vowel in advanced Polish learners of English: Beyond averages"

2:35-3:05 Isabela Lazar, "A morphosyntactic algorithm for sentence building in language acquisition"

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

A Case of Lexicalization: from Middle to Modern English (Chair: Anna Wärnsby)

3:35-4:05 Joanna Nykiel, "Do so and verb phrase ellipsis in the *Canterbury Tales*"

Section C (CAS 316)

Grammaticalization and degrammaticalization (Chair: Peter Siemund)

10:10-10:40 Julie Van Bogaert, "A multivariate analysis of that/zero alternation"

10:45-11:15 Marion Elenbaas, "Tracing grammaticalization in English light verbs"

11:20-11:50 Stefanie Wulff, "Gradient grammaticalization in English complement constructions"

11:55-12:25 Graeme Trousdale, "Ish"

12:30-1:55 LUNCH

Letters and Literature (Chair: Karin Axelsson)

2:00-2:30 Dustin Grue, "Relevance theory, accountabilities, and collocations in *Lord of the Flies* criticisms"

2:35-3:05 Minna Palander-Collin, "How can we study identity construction in early English letters?"

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Jim Walker, "The pre sent-perfect narrative in varieties of British English and farther afield"

Section D (CAS 324)

Workshop: Neal Norrick: Methods of Analyzing Spoken English

10:10-10:40 Neal Norrick, "Investigating Interjections in Narrative Contexts: A Hybrid Corpus Approach"

10:45-11:15 Gisle Andersen, "Corpus-driven approaches to discourse markers in spoken data"

11:20-11:50 Dagmar Barth-Weingarten, "The participants' perspective in interactional-linguistic work on the phonetics of talk-in-interaction"

11:55-12:25 Bruce Fraser, "Studying DM Sequences in Spoken English"

12:30-1:55 LUNCH

2:00-2:30 Christoph Rühlemann, “Introducing collocation analysis”

2:35-3:05 Klaus P. Schneider, “Just how useless are questionnaires for studying spoken language? Triangulating elicited and natural corpus data”

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Anne Wichmann and Nicole Dehé, “Corpus data and prosodic analysis”

4:10-4:40 Capstone Session

4:45-5:30 POSTER SESSION 1 (CAS 227) (Chair: Eugene Green)

Zeltia Blanco-Suárez, “Death-related intensifiers: Grammaticalization and related phenomena in the development of the intensifier *deadly*”

Daniele Franceschi, “Shall we *start* or ... *commence*? Stylistic aspects of near-synonymous verb use”

Mark Lindsay and Mark Aronoff, “Natural selection in self-organizing morphological systems”

7:00-9:00 CONFERENCE DINNER (Faculty Dining Room, 5th floor of George Sherman Union, 775 Commonwealth Avenue)

Speaker: Laurence Horn, Yale: “Etymythology and Taboo” (Introduction: Bruce Fraser)

MONDAY, JUNE 20

8:30-9:30 PLENARY SESSION (CAS 224)

Lisa Green “Multiple Grammars and Dialectal Variation: A View from the Perspective of Language Development” (Chair: Geoffrey Russom)

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION IV

Section A (CAS 213)

Early English constructions (Chair: Ilse Depraetere)

10:45-11:15 Ayumi Miura, “Lexical semantics in Middle English impersonal constructions”

11:20-11:50 Lieselotte Brems, “Fear(s) + complement clauses”

11:55-12:25 Izabela Czerniak, “Tracing the Scandinavian influence in early English”

12:30-1:55 LUNCH

Psychological aspects of English syntax (Chair: Izabela Lazar)

2:00-2:30 Carlos Prado-Alonso, “A cognitive approach to obligatory subject-

dependent XVS constructions in English”

2:35-3:05 Ute Römer, Matthew O'Donnell, and Nick Ellis, “Learning verb-argument constructions: New perspectives from corpus and psycholinguistic analyses”

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Rainer Schulze, “Aspects of seriality in language”

4:10-4:40 Laurel Smith Stvan, “The influence of lexical conflation on causation”

Section B (CAS 237)

Workshop: Hubert Cuyckens and Martin Hilpert: How can new corpus-based techniques advance historical description and linguistic theory?

10:10-10:40 Hubert Cuyckens & Martin Hilpert, “Introduction: How can new corpus-based techniques advance historical description and linguistic theory?”

10:45-11:15 Britta Mondorf, “Leg it, floor it, snuff it: A synchronic and diachronic analysis of nonreferential it”

11:20-11:50 Tanja Säily, “Sociolinguistic variation in morphological productivity in the CEECE”

11:55-12:25 Javier Perez-Guerra, “Pairing word order with headedness in the recent history of English: a corpus-based analysis”

12:30-1:55 LUNCH

2:00-2:35 Stefan Gries & Martin Hilpert, “Modeling diachronic change in a morpho-phonemic alternation”

2:40-3:05 Benedikt Szmrecsanyi, “Culture change versus grammar change: the limits of text frequency (and what we can do about it)”

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Maria José Lopez-Couso, “Corpus-based methodology and grammaticalization theory: Observing, describing, and analyzing grammaticalization and related processes of language change through corpus linguistics”

4:10-4:40 Workshop overview

Section C (CAS 316)

Workshop: Markku Filppula, Juhani Klemola, and Anna Mauranen: Global English: contact-linguistic, typological, and second-language acquisition perspective

10:10-10:40 Markku Filppula, Juhani Klemola, and Anna Mauranen: “Global English: contact-linguistic, typological, and second-language acquisition perspective”

10:45-11:15 Peter Siemund, “Varieties of English and Language Typology”

11:20-11:50 Niina Hynninen and Henrik Hakala, “Lexical and accent accommodation in ELF interaction”

11:55-12:25 Heli Paulasto, Elina Ranta, and Lea Meriläinen, “Syntactic features in Global Englishes: how ‘global’ are they?”

12:30-1:55 LUNCH

2:00-2:30 Edgar Schneider, “Tracking down American impact on Asian and Pacific Englishes in electronic corpora”

2:35-3:05 Hanna Parviainen, “Question formation in Indian English and in other Southeast Asian varieties”

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35-4:05 Zhiming Bao, “Systemic Nature of Substratum Transfer: the case of *got* in Singapore English”

4:10-4:40 Rajend Mesthrie, “Diamonds, gender and strong verbs: a study of contact and sociolinguistic factors in the evolution of a variety of Black English in Kimberley, South Africa”

4:45-5:30 POSTER SESSION 2 (CAS 314) (Chair: Eugene Green)

Jakob R. E. Leimgruber and Lavanya Sankaran, “Imperfectives in Singapore English: New evidence for ethnic varieties?”

Nadja Nesselhauf, “Diachronic corpus linguistics: overcoming the limitations of automatic analysis”

Carla Suhr, “Introducing visuals to historical pragmatics: Book history and multimodality”

5:35-6:35 PLENARY SESSION (CAS 224)

Stefan Gries, “The quantitative revolution in corpus linguistics: applications and their theoretical implications” (Chair: Martin Hilpert)

7:00-10:30 HARBOR CRUISE (Buses depart from 725 Commonwealth Avenue at 7)

TUESDAY, JUNE 21

8:30-9:30 BUSINESS MEETING (CAS 522)

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION V

Section A (CAS 213)

Modern English Constructions (Chair: Marion Elenbaas)

10:10-10:40 Ilse Depraetere and Chad Langford, “On the meaning(s) of *need to*”

11:20-11:50 Doris Schoenefeld, “Modern Usage and semantic change”

Section B (CAS 237)

Asian and Pacific English (Chair: Edgar Schneider)

10:10-10:40 Tatiana Larina, Svetlana Kurtes, and Neelakshi Suryanarayan,
“*Madam* or *auntie jee*: contrasting forms of address in British and Indian English(es)”

10:45-11:15 Manfred Sailer, “Doubling in New Englishes”

Orthographic Developments (Chair: Edgar Schneider)

11:25- 11:55 Christina Sanchez-Stockhammer, “Tracing orthographic change in corpora: A methodological approach to the study of English compound spelling”

Section C (CAS 316)

Contact or Comparisons of English and related Germanic languages (Chair: Gisle Andersen)

10:10-10:40 Anna Wärensby, “Interpreting modal utterances in English and Swedish”

10:45-11:15 Eline Zenner, “The borrowability of English”

Section D (CAS 324)

Developments of Idiosyncratic Constructions (Chair: Rainer Schulze)

10:10-10:40 Laurel Brinton, “The extremes of insubordination: exclamatory /as if!/?”

10:45-11:15 Beate Hampe, “A study of expressive *a(n) N of a(n) N* constructions in the BNC”

11:20-11:50 Georg Maier, “Pronoun case variation across varieties of English”

1:30-4:00 ARCHITECTURAL TOUR: DOWNTOWN BOSTON (Bus departs from 725 Commonwealth Avenue at 1:30)

Chairs

Bas Aarts, University of London

Gisle Andersen, Norwegian School of Economics and Business Administration

Karin Axelsson, University of Gothenburg

Markus Bieswanger, University of Flensburg

Laurel Brinton, University of British Columbia

Lynn Clark, University of Lancaster

Ilse Depraetere, University of Lille

Dagmar Deuber, University of Muenster

Stefan Diemer University of Saarlandes

Daniel Donoghue, Harvard University

Katie Drager, University of Hawaii

Marion Elenbaas, University of Leiden

Bruce Fraser, Boston University

Eugene Green, Boston University

Lauren Hall-Lew, University of Edinburgh

Stephen Harris, University of Massachusetts, Amherst
Martin Hilpert, University of Freiburg
Magnus Huber, University of Giessen
Gunther Kaltenback, University of Vienna
Izabela Lazar, University of British Columbia
Charles Meyer, University of Massachusetts
Heli Paulasto, University of Eastern Finland
Marnie Reed, Boston University
Geoffrey Russom, Brown University
Edgar Schneider, University of Regensburg
Rainer Schulze, University of Hannover
Peter Siemund, University of Hamburg
Elizabeth Traugott, Stanford University
Anna Wärensby, Malmö University

Kind thanks to all chairs, especially those not presenting.

ISLE 2011 Schedule Tables

THURSDAY, JUNE 16

4:00-8:00 REGISTRATION (CAS 116)

8:00-10:00 RECEPTION: Boston University Pub (225 Bay State Road, beneath BU Castle)

FRIDAY JUNE 17

8:00-9:00 REGISTRATION (CAS 116)

9:00-9:15 OPENING REMARKS (CAS 224)

9:15-10:15 PLENARY SESSION (CAS 224)

Christian Mair, "World Non-Standard Englishes: Reflections on the Global Spread of (Some) Vernacular Varieties of English"

10:15-10:45 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION I

	SECTION A (CAS 203)	SECTION B (CAS 237)	SECTION C (CAS 316)	SECTION D (CAS 226)
	<i>INTERNET IDIOMS</i>	<i>COMPARATIVE CONSTRUCTIONS</i>	<i>IRISH ENGLISH</i>	<i>VARIATIONISM</i>
10:50 11:20	JON BAKOS	THOMAS HOFFMAN	JULIA DAVYDOVA	DON CHAPMAN
11:25 11:55	KERREMANS, STEGMAYR	GUNTHER KALTENBÖCK	MARIJE VAN HATTUM	KIRK HAZEN
12:00 12:30	URSULA KIRSTEN	TURO VARTIAINEN	STEPHEN LUCEK	SANDRA JANSEN

12:30-1:55 LUNCH

	<i>CASE</i>	<i>AFRICAN/RELATED DIASPORIC ENGLISHES</i>	<i>PHONOLOGICAL TOPICS IN AMERICAN/NEW ENGLISHES</i>	<i>ACADEMIC STYLES</i>
2:00 2:30	PAYNE, BERLAGE	LARS HINRICHS	DAVID EDDINGTON	PRESLEY IFUKOR
2:35 3:05	WOLK, BRESNAN, ROSENBACH, SZMRECSANYI	HUBER, SCHMIDT	CAROLINE WILTSHIRE	UTE RÖMER

3:10 3:40	WULFF, GRIES	ROBERT FUCHS	TOSHIHIRO ODA	PETER SIEMUND
----------------------------	---------------------	---------------------	----------------------	----------------------

3:45-4:10 COFFEE BREAK (CAS 2nd floor hall)

	<i>PERCEPTUAL DIALECTOLOGY</i>			
4:15 4:45	CHRIS MONTGOMERY	GLENDALICIA LEUNG		

4:55 PLENARY SESSION (CAS 224)

David Denison, Presidential Address

SATURDAY, JUNE 18

8:30-9:30 PLENARY SESSION (CAS 224)

April McMahon, “Comparing [lark] with [lɜrk]: Methods for Collecting and Comparing Data from Varieties of English”

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION II

	SECTION A (CAS 213)	SECTION B (CAS 237)	SECTION C (CAS 316)	SECTION D (CAS 324)
	<i>WORKSHOP: MERGERS IN ENGLISH</i>	<i>MODERN ENGLISH CONSTRUCTIONS</i>	<i>WORKSHOP: GENITIVE VARIATION IN ENGLISH</i>	<i>WORKSHOP: ENGLISH IN THE INDIAN DIASPORA</i>
10:10 10:40	MAGUIRE, CLARK, WATSON	AARTS, BOWIE, WALLIS	PAYNE, BERLAGE	DEUBER, LEUNG, LACOSTE
10:45 11:15	MACIEJ BARANOWSKI	KARIN AXELSSON	TAGLIAMONTE, JANKOWSKI	MARIANNE HUNDT
11:20 11:50	CLARK, WATSON	LINNEA MICCIULLA	CATHY O’CONNOR	JAKOB LEIMGRUBER
11:55 12:25	DRAGER, HAY		BÖRJARS, DENISON, KRAJEWSKI	RAJEND MESTHRIE

12:30-1:55 LUNCH

		PRAGMATICS		
2:00 2:30	LAUREN HALL- LEW	MARKUS BIESWANGER	EHRET, WOLK, SZMRECSANYI	CLAUDIA RATHORE
2:35 3:05	JENNIFER NYCZ	CHRISTINE GÜNTHER	EVELIEN KEIZER	ALAM, STUART- SMITH

3:05-3:35 EXHIBITORS' COFFEE HOUR (CAS 216)

			CANADIAN ENGLISH	
3:40 4:10	PHILLIP TIPTON	MEIKE PFAFF	CHARLES BOBERG	LENA ZIPP
4:15 4:45		ALEXANDER BERGS	STEFAN DOLLINGER	CAPSTONE SESSION

4:55-5:55 PLENARY SESSION (CAS 224)
Christopher Ricks, “The Very Words, and Not Only Those”

SUNDAY, JUNE 19

8:30-9:30 PLENARY SESSION (CAS 224)
**Sali Tagliamonte, University of Toronto: “System and Society in the Evolution of
Change: The Case of Canada”**

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION III

	SECTION A (CAS 213)	SECTION B (CAS 237)	SECTION C (CAS 316)	SECTION D (CAS 324)
	CORPUS STUDIES	WORKSHOP: ASPECTS OF METHODOLOGY AND PEDAGOGY	GRAMMATICAL- IZATION AND DEGRAMMA- TICALIZATION	WORKSHOP: METHODS OF ANALYZING SPOKEN ENGLISH
10:10 10:40	GARRISON BICKERSTAFF	WALT WOLFRAM	JULIE VAN BOGAERT	NEAL NORRICK
10:45 11:15	TERTTU NEVALAINEN	PUNELL, RAIMY, SALMONS	MARION ELENBAAS	GISLE ANDERSEN
11:20 11:50	MATTHEW O'DONNELL	BILL KRETZSCHMAR	STEFANIE WULFF	DAGMAR BARTH- WEINGARTEN

	<i>CASE STUDIES</i>			
11:55 12:25	LIEVEN VANDELANOTTE	KIRK HAZEN	GRAEME TROUSDALE	BRUCE FRASER

12:30-1:55 LUNCH

		<i>WORKSHOP: EVALUATION AND INSTRUCTION</i>	<i>LETTERS AND LITERATURE</i>	
2:00 2:30	GREGORY GARRETSON	JAROSALAW WECKWERTH	DUSTIN GRUE	CHRISTOPH RÜHLEMANN
2:35 3:05	STEFAN DIEMER	IZABELA LAZAR	MINNA PALANDER- COLLIN	KLAUS SCHNEIDER

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

		<i>A CASE OF LEXICALIZATION: FROM MIDDLE TO MODERN ENGLISH</i>		
3:30 4:05	MICHAEL ERLEWINE	JOANNA NYKIEL	JIM WALKER	ANNE WICHMANN, NICOLE DEHÉ
4:10 4:40				CAPSTONE SESSION

4:55-5:30 POSTER SESSION 1 (CAS 227)

Zeltia Blanco-Suárez, “Death-related intensifiers: Grammaticalization and related phenomena in the development of the intensifier *deadly*”

Daniele Franceschi, “Shall we *start* or ... *commence*? Stylistic aspects of near-synonymous verb use”

Mark Lindsay and Mark Aronoff, “Natural selection in self-organizing morphological systems”

7:00-9:00 CONFERENCE DINNER (Faculty Dining Room, 5th floor George Sherman Union, 775 Commonwealth Avenue)

Laurence Horn, Yale, “Etymythology and Taboo”

MONDAY, JUNE 20

8:30-9:30 PLENARY SESSION (CAS 224)

Lisa Green, "Multiple Grammars and Dialectical Variation: A View from the Perspective of Language Development"

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION IV

	SECTION A (CAS 213)	SECTION B (CAS 237)	SECTION C (CAS 316)
	<i>EARLY ENGLISH CONSTRUCTIONS</i>	<i>WORKSHOP: NEW CORPUS-BASED TECHNIQUES</i>	<i>WORKSHOP: GLOBAL ENGLISH</i>
10:10 10:40		CUYCKENS, HILPERT	FILPPULA, KLEMOLA, MAURANEN
10:45 11:15	AYUMI MIURA	BRITTA MONDORF	PETER SIEMUND
11:20 11:50	LIESELOTTE BREMS	TANJA SÄILY	HYNNINEN, HAKALA
11:55 12:25	IZABELA CZERNIAK	JAVIER PEREZ-GUERRA	PAULASTO, RANTA, MERILÄINEN

12:30-1:55 LUNCH

	<i>PSYCHOLOGICAL ASPECTS OF ENGLISH SYNTAX</i>		
2:00 2:30	CARLOS PRADO-ALONSO	GRIES, HILPERT	EDGAR SCHNEIDER
2:35 3:05	RÖMER, O'DONNELL, ELLIS	BENEDIKT SZMRECSANYI	HANNA PARVIAINEN

3:05-3:30 COFFEE BREAK (CAS 2nd floor hall)

3:35 4:05	RAINER SCHULZE	MARIA JOSÉ LOPEZ- COUSO	ZHIMING BAO
4:10 4:40	LAUREL SMITH STVAN	WORKSHOP OVERVIEW	RAJEND MESTHRIE

4:55-5:30 POSTER SESSION 2 (CAS 314)

Jakob R. E. Leimgruber and Lavanya Sankaran, “Imperfectives in Singapore English: New evidence for ethnic varieties?”

Nadja Nesselhauf, “Diachronic corpus linguistics: overcoming the limitations of automatic analysis”

Martin Schweinberger, “A fine-grained variationist analysis of discourse marker LIKE across and within regional varieties of English”

Carla Suhr, “Introducing visuals to historical pragmatics: Book history and multimodality”

5:35-6:35 PLENARY SESSION (CAS 224)

Stefan Gries, “The Quantitative Revolution in Corpus Linguistics: Applications and Their Theoretical Implications”

7:00-11:00 HARBOR CRUISE (Buses leave from 725 Commonwealth Avenue)

TUESDAY, JUNE 21

8:30-9:30 BUSINESS MEETING (CAS 522)

9:35-10:05 COFFEE BREAK (CAS 2nd floor hall)

GENERAL SESSION V

	SECTION A (CAS 213)	SECTION B (CAS 237)	SECTION C (CAS 316)	SECTION D (CAS 324)
	<i>MODERN ENGLISH CONSTRUCTIONS</i>	<i>ASIAN AND PACIFIC ENGLISH</i>	<i>CONTACT OR COMPARISONS OF ENGLISH AND GERMANIC LANGUAGES</i>	<i>DEVELOPMENTS OF IDIOSYNCRATIC CONSTRUCTIONS</i>
10:10 10:40	DEPRAETERE, LANGFORD	LARINA, KURTES, SURYANARAYAN	ANNA WÄRNSBY	LAUREL BRINTON
10:45 11:15	DORIS SCHOENEFELD	MANFRED SAILER	ELINE ZENNER	BEATE HAMPE
		<i>ORTHOGRAPHIC DEVELOPMENTS</i>		
11:20 11:50		CHRISTINA SANCHEZ-STOCKHAMMER		GEORG MAIER

1:00-4:00 ARCHITECTURAL TOUR OF DOWNTOWN BOSTON (Bus leaves from 725 Commonwealth Avenue)

Conference Abstracts

Plenary Speakers

World non-standard Englishes: Reflections on the global spread of (some) non-standard varieties of English Christian Mair (University of Freiburg)

Globalisation has helped the spread and further entrenchment of Standard English in many obvious ways. What is discussed less often is the fact that globalisation has helped the spread of other languages, and of selected non-standard varieties of English, too. After a brief introduction, in which I will discuss the role of English in a globalising and multilingual world, I will trace the transnational impact of selected non-standard varieties of English, exemplifying the phenomenon mainly with data from Jamaican Creole, but casting an additional look into the direction of New Englishes from West Africa and India. It will emerge that, expectedly, non-standard varieties of English spread in the wake of global currents of migration but, probably less expectedly, the participatory media of the Internet age have assumed a crucial additional role in the process, too.

In my view, this calls for a re-assessment of some time-honoured assumptions in sociolinguistics/ EWL studies, such as the primacy of the „authentic“ locally based community vernacular or the central status of the notion of „variety of English“.

As I shall argue, many of the concerns raised in the paper can be addressed by conducting the study of varieties of English, World Englishes or New Englishes in the framework of the „sociolinguistics of mobile resources“ recently proposed by Blommaert to account for new types of language use, language spread and language contact in a globalising and multilingual world.

ISLE Highlights?

David Denison (University of Manchester)

Consider these recent examples of student writing:

- (1) this **highlights** once more **that** [...]
- (2) Sweet **defines that** “grammar may be regarded either from a theoretical or practical point of view. [...]
- (3) This study has **displayed that** older participants have more stable and confident results than [...]
- (4) Steven Pinker, (1994) **puts forward that** chimps often just imitate the messages of the trainer
- (5) Milroy et al [...] **utters that**, “In other locations [...]

Verbs which have traditionally taken nominal objects, including nominal objects with an embedded *that*-clause complement (*highlights the fact that ...*), are sporadically being used with a *that*-clause complement of their own. Is there a clear boundary between verbs like *highlight* and such factual and suasive verbs as *acknowledge*, *point out*, *recommend*, *suggest*, etc.? What might seem like mistakes can equally be regarded as evidence of change in progress.

I trace the development of these analogical patterns in corpus data and offer an analysis. A surprising number of methodological and theoretical approaches can be brought to bear on the problem: English linguistics is a rich and varied discipline. However, other questions are raised by data like this. Is there a conflict between our roles as teachers and researchers? Does prescriptivism have anything to offer here? What are the boundaries of English, or which Englishes should we be looking at? And so I come to more general reflections on the state of the language, the state of English linguistics, the teaching of our subject, and the state of our Society.

Comparing [laɪk] with [lʌɪk]:
Methods for Collecting and Comparing Data from Varieties of English
April McMahon (University of Edinburgh)

The relationship between language variation and language change is an area of considerable current debate. It is increasingly evident that divergence through time from a common ancestor, and contact between speakers and hence between languages, are relevant for both language variation and change. Arguably, variation and change are best considered as a continuum, and each can bring insights to the study of the other. Nonetheless, in some respects approaches to research in linguistic change and variation are growing apart – a trend perhaps encouraged by the models and metaphors which have been developed to capture language families on the one hand, and linguistic variation on the other. Families are typically shown as belonging to family trees, in which it is not possible to accommodate the results of contact, or the existence of common, independent but parallel changes. On the other hand, trees typically stop before we reach the level of accents and dialects, which are often considered within a wave model of mutual influence and continued contact.

Fortunately, it is also possible to find a more neutral and flexible means of representation, involving networks. Quantitative approaches first establish a way of measuring or counting language features; then pass the results through a program such as NeighborNet which show similarity regardless of its source or cause. In this talk, I shall show that such networks help us to move beyond the limitations of family trees, without losing sight of the real processes and outcomes of descent with modification from a common ancestor, but also without sidelining or ignoring variation. Using results from the recent Sound Comparisons project (see <http://www.soundcomparisons.com>), I shall show that we are able to use quantitative methods and network representations for the whole continuum of comparisons – between languages; between accents; between individual speakers; and between different historical stages of the same language. We therefore have a clear way of reintegrating variation and change.

However, these illustrations involve only articulatory, segmental phonetics, and are based on a single means of comparison and a single mode of displaying the results. We shall end with a more general overview of future needs for the field, including the development of new methods of quantification and representation across different levels of the grammar; an exploration of approaches to collecting data; and improved ways of making such comparative data more generally available. These emerging trends in collaborative working may in time transform our understanding of both variation and change.

“The very words, and not only those”:
Christopher Ricks (Boston University)

Christopher Ricks, who is not a linguist, will present several instances of how editing, whether of verse or prose, can sustain an appreciation of the English language, with the scholarly enterprise (textual and contextual) renewing an awareness of linguistic history – and of the relation of English to other languages. He is the author and editor of many books, including *Paradise Lost & Paradise Regained* (ed. 1968), *The Poems of Tennyson* (ed. 1987), *The Expelled / The Calmative / The End / First Love* by Samuel Beckett (ed. 2009) *What Maisie Knew* by Henry James (ed. 2010), *The State of the Language* (ed., with Leonard Michaels, 1980), and “American English and the inherently transitory,” *The Force of Poetry* (1984)

System and society in the evolution of Change: The view from Canada

Sali A. Tagliamonte (University of Toronto)

In this presentation I synthesize the results arising from a research program studying the transmission and diffusion of linguistic change in Ontario, Canada's largest province. Two contrasting linguistic features, each from distinct levels of grammar encapsulate the emerging findings: (1) stative possession and (2) quotatives.

(1) He *has* a fishing boat but it's *got* music in it.

(2) And he *said*, "What are you insane?" I'm *like*, "What does insane mean?"

The trajectories of change for these variables in Toronto, the largest city in Canada (Tagliamonte & D'Arcy 2007; Tagliamonte, D'Arcy and Jankowski 2010) provide a base-line. Although Canadian English is often thought to be one large dialect from sea to sea (Labov, Ash, & Boberg 2006); this homogeneity is an urban phenomenon (Chambers 1991:93). In recent years, fieldwork in cities and towns outlying the city are beginning to unearth a more nuanced picture of these developments (e.g. Tagliamonte & Denis 2008).

The data come from a small city and a smaller town in the southeast, one 2 and the other 4 hours away from the city (Belleville and Lakefield); and two small cities due north, one 300 and the other 800 kilometers away (Porcupine and North Bay). How are the changes in Toronto in (1-2) proceeding in these communities?

According to the wave model, linguistic change should spread outward from an originating centre while the cascade model predicts that change should diffuse to larger centers before small ones (Trudgill 1974). Thus, we might hypothesize that those distant from the city will provide a view of an earlier stage in the history of the changes in (1-2) and that the larger communities will be furthest ahead while the smaller, remote towns will lag behind. However, the progression of change also depends on how the change is diffused and by whom as well as its linguistic complexity (Kerswill 1996; Labov 2007). Longitudinal grammatical changes such as the development of possessive *have* can be expected to exhibit parallel developments in core vs. peripheral contexts through normal processes of transmission in each locale. However, a recent innovation such as quotative *be like* can be expected to be the product of diffusion.

Early results from southern Ontario (Tagliamonte & Denis 2008) revealed remarkable constancy across core and peripheral areas for both these features; however, with the addition of the two northern cities a more nuanced picture is emerging. The frequency of incoming forms, whether *have* or *be like*, are progressing in tandem regardless of local; however the constraints within each system are differentiated by community. In southern Ontario where British ancestry is predominant, the full gamut of older constraints is present. Far afield in the north, however, where there were mixed ethnic founders from the outset and a distinct economic, social and cultural base, these same constraints have been lost or weakened. The results show that both the system of the language as well as its socio-cultural embedding are crucial for interpreting and understanding language change.

In order to fully understand and interpret these differences the changes must be interpreted within the historical and socio-cultural context in which the change has developed.

Etymythology and Taboo

Laurence R. Horn (Yale University)

100 years ago, Freud found support for his edict that there is no “no” in the unconscious in Carl Abel’s “universal phenomenon” of *Gegensinn*, the antithetical sense of primal words purportedly abounding in the “oldest” languages. While time has not been kind to Abel’s and Freud’s thesis, philologists have long both recognized and argued over the general tendency for words in a wide variety of languages to develop and maintain contradictory or opposite meanings. Standard English examples include nouns like *sanction* or *oversight*, verbs like *cleave* or *dust*, adjectives like *bad*, and adverbs like *literally*. To a large extent, the persistence of such words is part of the more general phenomenon of homonymy tolerance: when and why do languages tolerate homonyms in the face of the general “Avoid Homonymy” principle and to the despair of prescriptive grammarians? The standard view is that it is only when the homonyms share the same categories and contexts of occurrence and a similar semantic field and register that one of them—the “weaker” homonym or near-homonym—will disappear (*strait* ‘narrow’, *let* ‘prevent’, *pale* ‘shovel’, *neer* ‘kidney’). But in the case of taboo words, as is well known (cf. Bloomfield 1933), Avoid Homonymy is generalized to block word senses or uses even when no confusion would plausibly occur, whence the suppression of *cock* ‘rooster’, *ass* ‘donkey’, *coney* ‘rabbit’, and more recently *niggardly* ‘stingy’. In this linguistic correlate of Gresham’s Law, “bad” (taboo) meanings force out “good” (innocent) ones; this presentation will survey the linguistic landscape of taboo avoidance and its role in word loss and meaning change. The power of taboo in language is also on display in the psychology of folk etymology, particularly in the domain of etymythology—folk etymology as lexical urban legend, revealing more about the language users than the language used. It is here that a legitimate argument can be made for a Freudian, or at least psychoanalytically informed, philology. One case in point will be examined in some detail: the *spittin’ image* (*spit and image*).

Biosketch

Laurence Horn was educated at the University of Rochester (BA 1965) and UCLA (PhD 1972 under the supervision of Barbara Partee). He is Professor of Linguistics and Philosophy at Yale University, where he has taught since 1981. He is the author of *A Natural History of Negation* (1989, reissue edition 2001) and numerous monographs, papers, and encyclopedia entries on negation, pragmatics, and word meaning. He also served as (co-)editor of *Negation and Polarity* (Oxford, 2000), *The Handbook of Pragmatics* (Blackwell, 2004), *Explorations in Pragmatics* (de Gruyter, 2007), and *The Expression of Negation* (de Gruyter, 2010).

**Multiple Grammars and Dialectal Variation:
A View from the Perspective of Language Development**
Lisa Green (University of Massachusetts Amherst)

Linguistic variation has been mostly in the domain of sociolinguistics, and work in that area has contributed significantly to the study of African American English (AAE), especially in addressing questions about linguistic variables in the context of social factors, linguistic constraints, and variation and change (e.g., Labov 1969, 1972; Rickford 1998; Poplack 2000). Much of the research on adult AAE has focused on the inherent variability of the linguistic system, and questions have been raised about the extent to which developing AAE-speaking children show signs of variation in the production of morphosyntactic features, such as the copula, that mirror the quantitative and statistical production reported for adult AAE. This paper builds on research in sociolinguistic variation and AAE and applies mechanisms in current syntactic theory related to nodes, structural positions of elements, and interpretable and uninterpretable features to empirical data to give an account of where variation occurs and how it is reflected in acquisition. I consider data from spontaneous speech and elicitation tasks as a means of analyzing variation in tense and aspect, negation, possessive, and agreement structures in child AAE in relation to the acquisition path. AAE exhibits syntactic and morphosyntactic variability, and it exhibits a significant amount of invariability, so it is an excellent case study for inspection of properties that account for where variation occurs and does not occur. This work has implications for claims about the locus of variation and the acquisition of multiple grammars.

Case

The effect of semantic relations on genitive variation

John Payne (University of Manchester)

Eva Berlage (University of Paderborn)

In this paper, using the British National Corpus (BNC), we present a multivariate analysis of the effect of semantic relations on the variation between *s*-genitive and *of* constructions in modern British English.

A common strategy in existing corpus studies of this variation is: (a) to identify a subset of examples in which variation is in principle possible, excluding all factors which categorically or near-categorically favour one construction or the other, e.g. personal pronouns as dependents (*her eyes* vs **the eyes of her*); then (b) to analyse this subset for the effect of factors such as animacy, weight, and the discourse status of head and dependent. The results of recent large-scale synchronic studies converge on the conclusion that animacy contributes most to the explained variation, with weight and discourse status also significant (O'Connor, Maling & Skarabela 2009; Börjars, Denison & Scott 2009).

However, as Gries (2002) and Stefanowitsch (2003) argue, the individual semantic relations involved between head and dependent are also a major factor in predicting the choice between the two constructions. Certain semantic relations must automatically be excluded from consideration in any multivariate analysis because they categorically favour the *of* construction (e.g. partitive *rest of the lecture*, container *basket of fruit*). But this leaves, we estimate, about twenty distinct semantic relations (at a relatively fine level of description) for which variation is permitted. In a recent diachronic corpus study, Szmrecsanyi (to appear) adds to a multivariate analysis a simple binary distinction between ownership and all other semantic relations. The fact that even such a coarse distinction leads to significance, coupled with Rosenbach's (e.g. 2002) experimental results based on a distinction between prototypical and non-prototypical relations, motivates the more fine-grained treatment in this paper.

Specifically, we select for each of a range of variation-permitting semantic relations a noun which functions as a typical head in that semantic relation, e.g. *eye* as a typical head in the body-part relation (*Mary's eyes* / *the eyes of Mary*), or *picture* in the descriptive theme relation (*Mary's picture* / *the picture of Mary*). The resulting data set, coded for semantic relations as well as factors such as animacy and weight, is then subjected to a multivariate analysis. This results in the establishment of a hierarchy of semantic relations according to their individual contributions to the choice between the two constructions.

Dative and genitive variability in Late ModE: Exploring cross-constructional variation and change

Christoph Wolk (Freiburg Institute for Advanced Studies)

Joan Bresnan (Stanford University)

Anette Rosenbach (Tanagra Wines)

Benedikt Szmrecsanyi (Freiburg Institute for Advanced Studies)

We present a novel, cross-constructional and corpus-linguistic approach to the analysis of language variation and change, drawing on richly annotated datasets and state-of-the-art multivariate analysis techniques. We are concerned with the history of the dative alternation (cf. Bresnan and Ford 2010), as in (1), and the genitive alternation (Rosenbach 2002, Hinrichs & Szmrecsanyi 2007), as in (2), in Late Modern English.

- (1) a. he deliver'd [my message] [to the King] <1749dodi.j3b>
(the prepositional dative construction)
- b. pay [his Royal Highness] [all the Honors due to his rank] <1793pen2.n4a>
(the ditransitive construction)
- (2) a. [the sanction] of [the KING] <1819mor1.n5b>
(the of-genitive)
- b. [the King]s [last Propositions] <1672lon2.n2b>
(the s-genitive)

Theoretically we follow sociolinguistic theory (e.g. Labov 1982, Tagliamonte 2001) and recent probabilistic accounts of language (e.g. Bod et al. 2003, Bresnan and Ford 2010) in assuming that syntactic variation – and change – is probabilistic rather than categorical in nature.

To this end, we tap ARCHER ("A Representative Corpus of Historical English Registers"), which covers the period between 1650 and 1990, spans 1.8 million words of running text, and samples eight different registers, British and American. Through substantial hand-coding, we derive extensively annotated datasets which characterize each dative or genitive observation ($N_{\text{total}} > 7,000$) in the corpus by way of a multitude of explanatory variables. Crucially, some of these are common to both alternations: consider the weight of the recipient/theme or possessor/possessum (cf. the principle of 'end weight'; Behaghel 1909), animacy of the recipient or possessor (cf. Rosenbach 2005), or definiteness of the recipient or possessor (cf. Bresnan, Cueni, Nikitina & Baayen 2007). The way these factors affect syntactic choices is not idiosyncratic for English but echoes cross-linguistic regularities. We also include a number of factors that are specific to each alternation, such as the semantic nature of the possessive relation in the case of the genitive alternation (cf. e.g. Taylor 1989). We then fit two logistic regression models with mixed effects (Baayen, Davidson & Bates 2008) that predict writers' dative and genitive choices by jointly considering all of the explanatory variables in the dataset while also allowing for idiolectal and lemma-specific random effects. The models we report have an excellent fit and correctly predict over 90% of the dative and genitive observations in ARCHER.

Overall, we find that the dative alternation is a bit more stable in real time – frequency-wise and also probabilistically – than the genitive alternation, which exhibits more variability. Nonetheless, we find some theoretically interesting communalities, such as the fact that the effect of more or less animate recipients in dative constructions and more or less animate possessors in genitive constructions appears to vary in lockstep. We interpret cross-constructional parallelisms like this against the cultural backdrop of, e.g., overall distributional changes in animacy categories, and we conjecture that such distributional fluctuations can trigger changes in probabilistic grammars in the long term. Thus, our probabilistic and cross-constructional approach to syntactic variation and change reveals general changes in grammar which might remain obscure when looking at these alternations in isolation.

A multifactorial study of the genitive alternation in L2 English

Stefanie Wulff (University of North Texas, Denton)

Stefan Th. Gries (University of California, Santa Barbara)

While alternations are arguably among the most intensively studied phenomena in native speaker English, research on their second language (L2) acquisition is surprisingly scarce. Little is known about which of the various factors associated with native speakers' choices between two variants learners pick up on, when they do so during the course of the L2 development, and how developmental paths vary as a function of their first language background.

This study seeks to take first steps towards closing this gap by investigating the L2 acquisition of the genitive alternation as illustrated in (1):

- (1) a. the squirrel'SPOSSESSOR nestPOSSESSEE
- b. the nestPOSSESSEE of the squirrelPOSSESSOR

Native speakers' choice between these two variants has been related to a variety of characteristics of the possessor and possessee noun phrases, such as their rhythmic alternation, segment alternation, morphological number marking, specificity, animacy, semantic class membership, syntactic branching, weight, and complexity, and activation status. (Selected relevant references are: Cooper & Ross 1975, Selkirk 1984, Hawkins 2004, Rosenbach 2002, Stefanowitsch 2003, and Hayes 2008.)

In order to determine if, and to what extent, intermediate-advanced German learners of English exhibit preferences similar to native speakers, we extracted all instances of the two genitive constructions from the German component of the *International Corpus of Learner English* (G-ICLE). A random sample of 1,000 attestations was hand-coded for all the variables mentioned above. Monofactorial statistics suggest that the German learners are well-attuned to the variables governing the genitive alternation: for each of the variables, chi-square tests confirm the learner data to be distributed significantly, and in the predicted direction when compared to the findings of native speaker studies.

More interestingly, a multifactorial analysis reveals a much more complex picture: a minimal adequate model of a binary logistic regression that includes interactions between variables yields a classification accuracy of 92.44%. Noteworthy deviations from native speaker studies will be discussed, such as the observed irrelevance of the noun phrases' activation status (a variable that was argued to be highly predictive of native speaker's constructional choices, if only in monofactorial analyses).

In order to shed more light on the question to what extent this overall striking match between German learners' and native English speakers' preference in the use of the genitive constructions is based on the fact that German learners benefit from positive transfer from their native to their second language, and in order to examine which variables are robustly associated with learners' constructional choices across different L1 backgrounds, we plan to report the results of a complementary analysis of more than 1,000 genitive attestations from Chinese learners (also obtained from ICLE) that we are currently carrying out.

Grammaticalization and Degrammaticalization

A multivariate analysis of *that*/zero alternation: A diachronic constructional study of the grammaticalization of *I think*, *I suppose* and *I believe*.

Julie Van Bogaert (University of California, Santa Barbara)

Christopher Shank (University of Leuven)

This paper examines the diachronic development of *that*/zero complementation alternation (1 – 2) with three complement-taking mental predicates (CTMPs), viz. *I think*, *I suppose* and *I believe*, and its relation to the emergence of these expressions as epistemic parentheticals (EPARs) (3 – 4). According to Thompson and Mulac's (1991) matrix clause hypothesis, increased use of the zero complementizer, as in (2), is a precondition for a CTMP's grammaticalization into an EPAR with adverb-like distribution, as in (3) and (4).

1. I think that the new settlement erm should be as close to York erm city as can be achieved. (BNC: HVK 69)
2. I think the County Council's behaviour over the Wheatley site is positively immoral. (BNC: KRL 3549)
3. And it's nice, I think, to include this plough, if that's what it is. (BNC: HM2 453)
4. Phil's already got a copy anyway. So's everybody else I think. (BNC: KLV 1160)

This paper takes its cue from Van Bogaert (2010), which presents a constructional taxonomy of CTMPs in which the high-frequency CTMP *I think* acts as a template onto which the other members of the taxonomy are modelled. Whereas Van Bogaert (2010) focuses on patterns of TAM variation, this paper investigates to what extent *I think* fulfils a similar template function when it comes to *that*/zero alternation. On the basis of findings put forward in Cuyckens and Shank (2009), we formulate the hypothesis that with regard to the spread of the zero complementation pattern too, *I think* acts as a 'pacemaker' (Van Bogaert 2010); due to the cognitive salience of this high-frequency CTMP and its analogy relationship with the rest of the taxonomy, the other CTMPs, which initially lag behind in this development, follow suit. Cuyckens and Shank (2009) diachronically study the *that*/zero alternation with *think*, *feel* and *realize* and find that the increase in zero use takes off earlier for *think* than for the other two CTMPs (cf. also Rissanen 1991). In the present paper, the low-frequency pair *feel* and *realize* is traded in for *I suppose* and *I believe*, which come second and third respectively in terms of frequency in the BNC (British National Corpus) (Van Bogaert 2010; to appear).

The study makes use of corpus data sets of equal size (50 tokens) representing eight different periods, from Early-Modern English (the mid-16th century) up to the present. Using logistic regression analysis, we examine the statistical significance of seven factors as to the selection of *that*/zero. It is hypothesized that the CTMPs display an increase in zero use which eventually reaches a plateau in Present-Day English. Once this threshold has been reached, the potential of zero use has been maxed out; there can be no further increase seeing that the alternative with explicit *that* is not precluded due to grammatical persistence (Torres Cacoullos & Walker 2009) and because it needs to remain available for processing reasons (cf. the impact of the factor 'length of the complement clause subject' on *that*/zero alternation) or to mark formality. As in Cuyckens and Shank (2009), *I think* is expected to come to this block first, as such setting an example to the other two CTMPs. Following Thompson and Mulac's (1991) main clause hypothesis, we expect to see a rise in the EPAR use of MPSs concomitant to the increase in zero use.

Tracing grammaticalisation in English light verbs: A corpus-based study
Marion Elenbaas (Leiden University)

This paper presents the results of a corpus study into the grammaticalisation of English Light Verbs (LVs), concentrating on Light Verb Constructions (LVCs) consisting of a LV and an indefinite deverbal nominal complement, e.g. *take a look*, *make a turn*, *have a smoke*. Brinton (2008) notes that such LVCs “have emerged primarily since the Later Middle English period” (45) and that “we must look to the Late Modern English period for the development of aspectual meaning ..., and such a study has yet to be done” (48). This paper looks at this Late Modern English period, as well as the (Late) Middle English and Early Modern English period (mostly using the *Penn Parsed Corpora of Historical English*). Some examples are presented in (1).

- (1) a. there he had a grete falle
‘There he had a big fall’ (cmmalory, 663.4718)
- b. in w=ch His Grace at first made an ingenuous confession.
(1673-1675, AUNGIER-E3-P2,170,A.18)
- c. My wife, my daughter, and herself, were taking a walk together.
(1766 O. Goldsmith, *Vicar of Wakefield*, II. ix. 137)

LVs give the impression of gradience: they appear to be somewhere in between lexical verbs (form-identity) and auxiliary verbs (lack of lexical meaning) and have indeed variously been considered to be lexical, semi-lexical or grammatical verbs. This paper adopts the view that apparent synchronic gradience is the result of fine categorial distinctions (see Roberts 2010), which arise from small-step changes. In order to elucidate (the changes in) the categorial status of LVs and reveal the extent of grammaticalisation, this paper closely examines the (aspectual) semantics and the morphosyntactic properties and possibilities (see e.g. the modification in (1a-b)) of LVs and LVCs. The Late Modern English data in particular, will be examined for signs of grammaticalisation of LVCs, such as the productivity of LVCs and host-class expansion of LVCs (see Brinton 2008).

Gradient grammaticalization in English complement constructions

Stefanie Wulff (University of North Texas, Denton)

In English subject-, object-, and adjectival complement constructions, the complementizer *that* is optional:

- (1) a. She thought (that) her colleague was in New York. (direct object)
b. The problem is (that) we don't have internet access here. (subject complement)
c. I'm glad (that) I found my purse later that night. (adjectival complement)

With regard to the question what determines the variable presence of *that*, various studies have pointed out that the absence of the complementizer is permissible with matrix clause verbs that denote truth claims (Dor 2005), as well as high frequency verbs in general (Tagliamonte & Smith 2005). Other studies observed that the complementizer is less often realized in spoken and informal discourse (Kaltenböck 2006). Moreover, there an effect of syntactic weight was attested such that light matrix and complement clause subjects (pronouns in particular) license omission of *that* (Elsness 1984). Another syntactic factor is clause juncture: *that* is more often omitted when no material intervenes between the matrix subject and the matrix verb, the matrix verb and *that*, or *that* and the complement clause, respectively (Rissanen 1991).

While there seems to be general agreement that the variable presence of *that* is best accounted for from a multifactorial perspective, most previous research presented monofactorial analyses, often based on selective samples of data (but see Tagliamonte & Smith 2005, Jaeger 2010). Also, no study so far has systematically compared the three kinds of complementation constructions in (1a-c). The present paper seeks to take a first step towards closing this gap by presenting a multifactorial analysis of all instances of the three types of complementation attested in the British component of the *International Corpus of English* (5,648 tokens).

The 5,648 corpus attestations were coded for a total of ten factors proposed in previous research, including the ones mentioned above. A logistic regression model suggests that disrupted clause juncture and heavy subjects and complements promote the production of *that*, especially in written discourse; the absence of *that* is associated with pronominal main and complement clause subjects and high frequency predicates ($p=0***$; $R^2=0.452$; $C=0.846$). These results are compatible with various processing accounts: *that* resolves temporary ambiguities (Bolinger 1972), shortens dependencies (Hawkins 2004), and aids in a maximally uniform distribution of information across the utterance (Jaeger 2010).

A closer look at the complementation type-specific verb associations (by means of a Distinctive Collexeme Analysis; Stefanowitsch & Gries 2005) furthermore suggests that the three complementation constructions are systematically associated with different sets of verbs, as are the respective instances within each complementation construction with or without *that*. DO-complements prefer verbs of knowledge assertion (such as *suggest*, *show*, *confirm*) when combined with *that*, while the verbs associated with DO complements without *that* tend to denote less speaker certainty (such as *think*, *suppose*, *guess*). While in ADJ-complementation, the semantics of the associated verbs overall shift from the speakers' degree of certainty of the validity of the proposition in the complement clause towards his/her emotional attitude towards the proposition in the complement clause, corresponding preferences can be observed depending on whether or not *that* is realized: *aware*, *clear*, and *satisfied* are significantly preferred when *that* is present, while *sure* and *afraid* prevail in contexts without *that*.

These findings invite a reconceptualization of the three complementation patterns as individual constructions in the sense of Goldbergian Construction Grammar (Goldberg 2006), occupying different locations on a validation-epistemic continuum. Likewise, the complementizer *that* is best seen as a construction in its own right (as opposed to a shallow grammatical surface structure). Moreover, the probabilistic nature of these complementation-(*that*)-verb preferences lends empirical support to gradient grammaticalization accounts (Thompson & Mulac 1991) taken into question in recent processing-focused accounts (Jaeger 2010), which, overall, constitutes further evidence in favor of the view that performance shapes grammar.

Ish

Graeme Trousdale (University of Edinburgh)

This paper is concerned with the historical evolution of the Old English (OE) suffix *-isc*, whose reflexes appear in a number of constructions in contemporary English. In some cases, the development suggests lexicalization (Brinton and Traugott 2005) of an erstwhile complex structure into an unanalysable referential whole (e.g. *English, Welsh, French*). In others, however, the development appears to be quite different, in that the suffix appears to have undergone degrammaticalization (specifically, debonding, as the term is used by Norde 2009) and can now be used as a clitic attaching not just to lexical categories such as nouns (*waspish, childish*) and adjectives (*greenish, coldish*), but phrases (*out-of-the-way-ish, devil-may-care-ish, rather-you-than-me-ish*), and even as stand-alone conversational turns (A: *Did you like it?* B: *Ish*). Taking evidence from a range of corpora (of contemporary and historical English), the paper considers specifically whether the development of *ish* is a clear-cut case of antigrammaticalization (Kuzmack 2007), and how this fits in with research on *ish* as a marker of vague language (Ruzaitė 2010). The story of *ish* seems to be rather idiosyncratic. Its development is unlike that of conversions of function words (cf. *to up the offer, to down a pint*) and that of the lexicalization of affixes (cf. *isms, ologies*), because the development of *ish* is a gradual process (unlike conversion), and does not result in the development of a hypernym (as in the case of *isms* and *ologies*). Furthermore, the history of *ish* raises important questions about the nature of (de)grammaticalization, since it involves the development of a derivational affix rather than an inflectional one, and the place of derivational morphology in (de)grammaticalization has been identified as contentious by many researchers (e.g. Hopper and Traugott 2003, Brinton and Traugott 2005, Norde 2009).

The paper traces individual developments (such as the derivation of denominal gentle adjectives, of depreciatory adjectives, and the evolution of free adjunct forms) within the framework of diachronic Construction Grammar, and particularly, within the framework of constructionalization (Traugott 2008, Fried 2008), the co-evolution of new form-meaning pairings which develop a grammatical or a lexical function. The story of *ish* is particularly interesting because it sits at the juncture of many different kinds of change (lexicalization, degrammaticalization, and pragmaticalization, for instance) which on the surface appear contradictory, but which may be unified within a constructionalization framework: some of the developments of *ish* suggest a case of lexical constructionalization with some of the form-meaning pairings (particularly those involving denominal gentle adjectives, and some of those involving denominal depreciatory adjectives), and grammatical constructionalization with others (for example, in cases where *ish* is linked to the growth of the degree modifier construction, Traugott 2008). However, these patterns become clear only when the development of individual (sets of) form-meaning pairings are considered as part of a reconfiguration of the English constructional network.

Modern English Constructions

On the meaning(s) of *need to*

Ilse Depraetere and Chad Langford (University of Lille)

Research on modal markers of non-epistemic necessity has focused especially on *must* and *have (got) to* (cf. e.g. Tregidgo 1982, Larreya 1982, Tagliamonte 2004, Depraetere and Verhulst 2007), with *need to* tacitly relegated to secondary status. It is interesting to note, however, that in his diachronic analysis of (semi-)modals of strong necessity, Smith (2003) observes that the use of *must* has declined and that it is *need to* rather than *have (got) to* that is gaining ground, effectively filling the gap left by *must*. While *have to* remains numerically the most common marker of necessity in the twentieth-century corpora of written British English, spoken British English and written American English that Smith analysed (cf. also e.g. Biber *et al* 1999: 489), the use of *need to* appears to have risen in these corpora by 266%, 650% and 123% respectively between 1960 and 1990. *Need to*, furthermore, has grown, in all types of syntactic environments. Smith argues that there is a considerable amount of semantic overlap with *must* and *have to*, the uniqueness of the marker being that it combines the notion of 'imposed obligation' (2003: 260) with that of 'required action (...) recommended for the doer's own sake.' This characterisation is in line with an observation in Leech (2003), who uses the label 'inner compulsion' to capture the kind of subject-internal source of the necessity typical of one use of *need to*.

The discussion of auxiliary *need* and lexical verb *need to* in both Quirk *et al* 1985 and Huddleston and Pullum *et al* 2002 is restricted to observations about the formal difference between auxiliary *need* and the lexical verb *need to*. Coates (1983) does not deal with *need to* and Palmer (1991) says that *need to* is 'roughly equivalent to *have to*' (1991: 128) although it 'does not simply have the necessity meaning of *must*, but rather indicates what is required for specific purposes or personal reasons'. This context of use is also mentioned in Declerck (1991) (*He needs to earn money if he is to buy a car*), who adds Leech's 'internal necessity' as a context typically requiring the use of *need to* (*She needs to talk to others about her work*). (1991: 386)

All in all, *need to* has received relatively little attention in previous research. It is also the object of little attention in pedagogical grammars, which devote greater space to the nonepistemic *must* vs. *have (got) to* distinction and mention *need to* as a kind of blanket synonym for both. It is this gap we aim to fill with our paper. We will offer an in-depth analysis of the meaning(s) and contexts of use of *need to* and will concentrate on how it competes with *have to*, its primary rival for the expression of necessity.

We will start with a quantitative and a qualitative analysis of *need to* in ICE-GB (340 tokens), and will then highlight the contexts of use of *need to*. A comparable number of tokens from the COCA corpus (1999-2009) will be analysed in order to check whether the quantitative differences between American and British English signalled by Smith (2003) are confirmed by these two recent corpora of English.

The analysis will reveal that it is not only syntactic, semantic and pragmatic considerations that lead the speaker to use this form, but that more often than is the case for modal markers like *must* and *have to*, it is lexical and collocational patterns that underlie the use of *need to*. To this end, we advance the hypothesis that, though easily blurred, there does exist a core distinction between the linguistic expression of necessity (via *have to*) and that of need (via *need to*) and that clear collocational patterns naturally fall out from this.

**Modern usage and semantic change:
The case of *go* and *come* plus participial complements**
Doris Schönefeld (University of Leipzig)

The talk will report on the results of an empirical study of particular uses of the English motion verbs *go* and *come*. In English reference grammars (such as Quirk et al. 1985), the reader will learn about constructions consisting of motion verbs and a following participle (present or past), with the latter completing the predication by adding information on the participant involved. It is assumed that in such constructions the primary meaning of the main verb (ie the motion verb) is weakened (cf ibid: 506).

For testing this claim empirically, we analysed the pattern(s) in modern English usage on the basis of data from the BNC. First, all forms of *go* in such a constructional environment were extracted and submitted to a collexeme analysis in order to see which participle forms are particularly attracted to this usage and to identify the meaning of the construction (for an argument regarding the performance of such analyses see Stefanowitsch and Gries, 2009: 943). For the determination of the meaning(s) of the *go* + participle pattern it turned out that the subjects needed to be incorporated as well. For, the subjects' in-/capability of self-propelled motion is strongly involved in triggering different senses which the constructions are meant to convey. The results of a co-varying collexeme analysis were helpful in further grouping the uses into various senses. These are predominantly depictive (i), and attributive readings (ii). For present participles, (ii) is not an option, instead a purposive sense and such of disapproval, inchoation and an ergative reading of an otherwise transitive scene (iii) can be identified.

- (i) ..., *staff must go accompanied to mental health assessments*; (BNC, GXJ)
- (ii) ..., *important posts in Dutch diplomacy had sometimes to be filled by second-rate men or go unfilled for longer periods*. (BNC, HY5)
- (iii) ..., *some of the files had gone missing years back*. (BNC, HDC)

The purposive and depictive readings are associated with the literal-motion sense of *go*, whereas the other readings imply a weakened concept of motion. The 'common denominator' in the semantics of these uses is the image-schematic structure associated with the verb *go*. The verb contributes the MOTION schema (SOURCE-PATH-GOAL) and the individual senses profile different parts of the schema and blank off others. Hence, the uses can be arranged in a network of senses extending from the basic meaning of literal motion. Further diversification can be accounted for by the two types of participles being involved.

An analogous analysis is being done for the *come* + participle pattern to see if the constructional diversity is (motion) verb specific or can be linked with the more general pattern of 'motion verb + participle'. Due to the differences in the conceptualizations of GO and COME, for example with respect to their deictic orientation, we assume the former assumption to be correct.

The final comparison of the two verbs in this pattern will be based on a distinctive collexeme analysis, which will help to identify the different construals triggered by the one or the other verb.

Typical and atypical change in modal usage over time

Bas Aarts, Jill Bowie and Sean Wallis (University College London)

It is frequently noted that modal auxiliary verbs are changing in frequency over time, but important and unexplored questions remain over the direction of travel and whether modals are alternating with one another. Thus Millar (2009) reports an increase in modal use over the course of the 20th century in the *TIME Corpus*, whereas Leech (2003) identifies a fall in written British English in the latter part of the century. Both authors demonstrate that distinct modal verbs appear to behave differently.

In this paper we present data from the *Diachronic Corpus of Present Day Spoken English* (DCPSE) spanning the period 1960–1990. The corpus is subdivided into two approximately equal subcorpora by time, and we measure the percentage difference between the later and earlier parts. We calculate 95% confidence intervals on this measure. This permits us to easily visualise the observed changes, as shown in Figure 1.

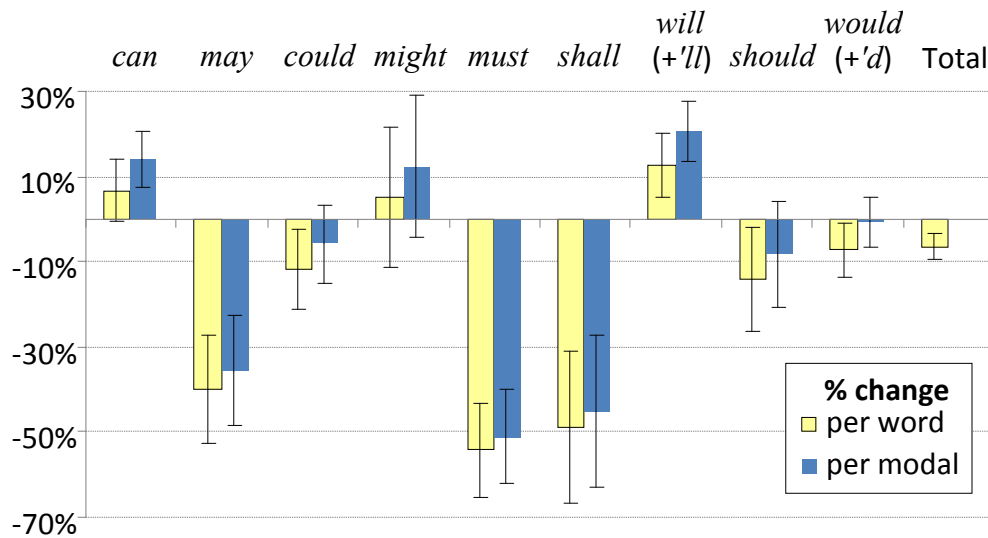


Figure 1: Percentage change from 1960s to 1990s in the modal auxiliary verbs.

This figure shows change relative to corpus size (in the ‘per word’ columns) and finds that the overall trend is downwards. In these absolute terms, the modal auxiliary verbs *could*, *may*, *must*, *shall*, *should* and *would* (including ‘d’) all fall in frequency, whereas *will* (including ‘ll’) is in the ascendant. We include negative cases and contracted forms in our data.

Second, we consider change *within* the data set (in the ‘per modal’ columns), i.e. relative to the overall downward trend we have already noted. We now find that, within the set of modals, *can* significantly increases relative to the overall trend (i.e. it ‘bucks the trend’). We refer to *could*, *would* and *should* as being ‘typical’ in that they undergo a significant decline, but this decline cannot be distinguished from the overall trend, whereas *can*, *may*, *must*, *shall* and *will* are ‘atypical’ as their changes significantly differ from the overall trend.

These findings suggest that among the set of modal auxiliaries some degree of change is explained by (partial) replacement of one form by another, for example of *shall* by *will* and of *may* by *can*. In this paper we will explore whether the trends observed are explained by changes in the semantic distribution (e.g. root versus epistemic) of the forms involved. Our findings will have implications for a finer-grained characterization of recent change in modal usage in spoken English.

A new functional model for tag questions based on fiction dialogue data

Karin Axelsson (University of Gothenburg)

Tag questions are mainly used in real-life conversation, as in (1), but they are also found in fiction dialogue, as in (2) and (3):

- (1) That's quite a good one, isn't it? (BNC spoken demographic)
- (2) "It's not dangerous, is it, Anthony"? (BNC written)
- (3) "You know that for a fact, do you, ma'am?" (BNC written)

Corpus research on the functions of tag questions has so far mainly focused on real-life conversation (e.g. Holmes 1995, Tottie & Hoffmann 2006). The study this paper reports on shows that research with a point of departure in fiction dialogue may contribute to the understanding of the pragmatic functions of tag questions in general.

Fiction dialogue can be described as "the writer's attempt to portray everyday natural language conversation" (Oostdijk 1990:235); it has been stated that "many crucial structural and functional principles are at work just as much in fictional dialogue as in natural conversation. It is hard to see how we could recognize and respond to the former as a version of the latter if this were not so" (Toolan 1990:275). An advantage of studying tag questions in fiction is that there is a great deal of information in the reporting clauses and the surrounding narrative giving clues as to how the author intends the reader to interpret the functions of the tag questions in the dialogue; for example, there may be comments on what the speakers and the addressees know beforehand about the matters brought up in the tag questions as well as information on non-responses and extra-linguistic responses such as nods and headshakes.

In the present study, a hierarchical model for the functions of tag questions with declarative anchors (both reversed-polarity tag questions, as in (1) and (2), and constant-polarity tag questions, as in (3)) has been developed based on fiction dialogue data from the *British National Corpus* (BNC) and then applied also to conversational data from the spoken demographic part of the BNC. The main distinction in the model is made between response-eliciting and rhetorical uses of tag questions. Response-eliciting tag questions are divided into a confirmation-seeking category, where the speaker is genuinely uncertain, and a confirmation-demanding category, where the speaker is quite certain and demands a confirmative answer to be uttered; response-eliciting tag questions may also be used primarily to start a conversation. Rhetorical tag questions are divided into a speaker-centred category, where the speaker's convictions, assessments etc. are in focus, and an addressee-oriented category, where the addressee is crucial, as such tag questions somehow concern the addressee. Several sub-categories have also been identified in some of these categories.

By applying this functional model to both samples, it has, for example, been possible to show that most tag questions are used rhetorically, and that there are functional differences between reversed-polarity and constant-polarity tag questions.

Factors Predicting the Use of Passive Voice in Newspaper Headlines

Linnea Micciulla (Boston University)

Information packaging researchers have found that certain factors influence active/passive voice alternations: Animacy, Definiteness and Weight influence argument order and thus choice of voice. Researchers in Critical Discourse Analysis (CDA) and psycholinguistics claim that voice is influenced by social factors, e.g. gender, social standing, or political bias. This dissertation draws from these distinct perspectives to perform probabilistic analysis of factors predicting voice in newspaper headlines, a novel research area for information packaging, and a rich source of data relevant to CDA.

In the first study to examine the relative contributions of these two types of constraints, this dissertation explores the predictive values of Animacy, Definiteness and Weight, as well as four social constraints: Gender, Nationality, Age and "Badness." It also investigates using combined human and automated methods for quick and accurate data annotation. The corpus consists of US newspaper headlines published between 2002 and 2007 containing one of twelve selected verbs: *accuse, aid, anger, create, encourage, frustrate, hit, hurt, injure, inspire, kill* and *shoot*.

The Animacy, Definiteness and Weight hierarchies predict that animate arguments tend to precede inanimate arguments, definite arguments tend to precede less definite arguments, and shorter arguments tend to precede longer arguments, respectively (Quirk et al. 1972, Ransom 1979, *inter alia*). The present findings support these hierarchies. Of the linguistic factors, Animacy has the strongest effect. Of the social factors, Nationality and Age are not significant predictors of voice, while Badness is a significant predictor. A "Bad" argument has an increased likelihood of occurring post-verbally relative to other arguments, so that a "Bad" Actor predicts passive, while a "Bad" Undergoer predicts active voice. Gender has a marginally significant effect which differs by verb; overall, arguments with a Female Actor are likely to occur with active voice relative to Male Actors; when the verb is *kill*, Female Undergoers are relatively more likely to occur with active voice.

The findings indicate that both social factors and traditional linguistic constraints predict voice. The results show that including social factors improves probabilistic models of grammar, and that analyses which include both linguistic and social factors provide better support for empirical claims.

Early English Constructions

Lexical semantics of Middle English impersonal constructions:

Approaches to and beyond the data

Ayumi Miura (University of Manchester)

This paper is concerned with the syntax-semantics interface of Middle English verbs of emotion, with particular reference to the (im)possibility of their being employed in impersonal constructions, as in:

Na man es forboden..to trowe in what lawe þat him list leue on.

‘No man is forbidden ... to believe in whatever law it pleases him to believe in.’
[c1400 *Mandeville’s Trav.* (Roxb.) xxiii. 108; *OED* s.v. list, v¹. 1.a.]

Impersonal constructions in Old and Middle English have been extensively explored for more than a century, ranging from van der Gaaf (1904) to Möhlig-Falke (2009). However, up to this day no serious investigation has been conducted into the question of why some verbs participate in impersonal constructions (e.g. LIKE and LOATHE) while others do not (e.g. LOVE and HATE), even though their semantics look closely related (see however Denison 1990, Allen 1995). The lack of attention to this topic contrasts sharply with the situation in the linguistics of Present-day English, where the link between verb senses and syntax has seen lively discussions (cf. Kearns 2006).

This paper builds on the methodological approaches in Elmer (1983), Allen (1995) and Carroll (1997) as well as studies of psych-verbs in modern languages (e.g. Croft 1993, Arad 1999, Pykkänen 2000) and compares the behaviour of several groups of Middle English verbs of emotion. I will argue that causation and stativity, which are key factors in the psych-verb literature, are also relevant to the presence, absence and spread of impersonal usage with these verbs. For a few groups, argument alternation and animacy of the Cause have some effect as well.

Another issue that this paper will address is the choice of database in view of the fact that a number of the target verbs are fairly rare and prevent us from obtaining ample evidence from standard corpora. I intend to demonstrate that, in such a context, lexicographical evidence has real advantages over conventional corpus material and allows one to build interesting generalisations about the syntax and semantics of impersonal verbs. To support this argument I will make full and careful use of the information from the *MED* entries.

Fear(s) + complement clauses:
a source of complex subordinators (*for fear of*) and modal particles (*No fear*)
Lieselotte Brems (University of Leuven)

'Noun phrases + complement clauses' (NPs + CCs) are the source construction of at least 2 distinct grammaticalized uses in English: complex subordinators (e.g. *in case (of)*, *in the event of*) and modal particles (e.g. *(I have/there is) no question*) (Kjellmer 1998). In both cases there is a shift from lexically full uses of the incorporated semiotic noun towards the expression of more grammatical meanings, hence grammaticalization. In this presentation I will address NPs + CCs as an under-researched locus of variation and change by means of an in-depth diachronic corpus study of '(preposition) (determiner) (modifier) *fear(s)* + *that/of/zero*-complement clause'. Synchronic variation suggests that this pattern has engaged in both of the processes described above. Based on extensive data from the PPCME, PPCME, CLMETEV and *WordBanks Online corpus*, the aim of this presentation is to track the diachronic micro-developments of the synchronic layering of this pattern from 1150-present.

Synchronically, *for fear (of/that)* can function as a complex subordinator expressing (negative) purpose between two clauses, similar to *lest* or *so as not to*. It expresses that one action is intentionally not performed so as to prevent causing another (undesired) and as yet non-factual action:

(1) *Gareth refrained from putting on the powerful light **for fear of** waking Rick, who lay like a dead man* (Collins Wordbanks Online or CWO)

There is no fear, no fear(s) and *have no fear(s)* synchronically act as epistemic modal markers (2) and disjuncts (3) with scope over the entire utterance, expressing strong certainty:

(2) *I **have** absolutely **no fears** that France will cave in to the strikers' action* (CWO) ('they won't cave in')

(3) *When I asked him if he wanted to change his mind, William said '**No fear**'*. (CWO) ('certainly not')

By means of careful qualitative and quantitative corpus analysis, I will reconstruct constructional stages in the development from lexical *fear(s)*-strings into grammaticalized ones, specifying the semantico-structural and pragmatic properties of source and bridging contexts (Diewald 2006). The relative chronology of the original and extended uses will be teased apart. The following questions will be addressed in this analysis: how can the CC become reanalyzed as a subordinated clause and 'preposition + NP + (particle)' as subordinator? It seems that in the immediate source construction of this use the NP with CC functioned as an adjunct in the matrix clause. In modal particle uses too, the structural and (discourse-)semantic status of the CC is reanalyzed. In the case of modal disjunct uses such as (3) the CC is in some way structurally reanalyzed as a modalized proposition. For the emergence of both types of extended use I will take into account the sorts of CC (*to* vs. *that* cf. 'proposition' vs. 'desired actions'-contents of the CC, Wierzbicka 1988), changes in the modification patterns of *fear(s)* by means of determiners, adjectives as well as the potential influence of phraseologies, e.g. *have/there is* + *fear(s)*, (negative) polarity sensitivity, e.g. *no fear(s)* and specialization in terms of 'discourse contexts', e.g. types of subjects (impersonal *there*, first persons) and verb phrases in the matrix clause and in CCs.

By means of a methodology that combines insights from construction grammar, grammaticalization research and diachronic corpus analysis this presentation makes an explicit stand for combined methodologies in linguistic analysis.

**Parsed diachronic corpora and language contact studies:
Tracing the Scandinavian influence in early English**
Izabela Czerniak (University of Eastern Finland, Joensuu)

The rise and stabilisation of the SVO order at the early stage of development of English is assumed to have occurred ultimately as a result of both internal and external language pressures. Among the latter, medieval cross encounters between the Anglian and Scandinavian populations are mentioned as a potential trigger for further morpho-syntactic reanalysis. The inflectional erosion, attributed to this particular contact situation (cf. Iglesias-Rabade), especially the petering out of cases on noun phrases, could, indeed, further reinforce the shift in word order, once the older manner of indicating syntactic functions had been disturbed (cf. Robinson 1992: 166).

Studies on medieval English texts have already pointed to the northern dialect as the most advanced in the inflectional attrition (e.g. Milroy 1992; Fischer 1992). The results of my own searches through two electronic corpora of Old (YCOE) and Middle (PPCME2) English have likewise placed the North ahead of other dialect sets with marked preference for SVO. In fact, not only did the loss of inflections and the overt stabilisation of SVO converge in one location but the latter progressed more rapidly throughout the areas affected by the contact (Czerniak 2009). The current study reveals that virtually any modification to the database output, whether on account of the normalisation of time frames or due to the exclusion of the translated material, does not alter the course of the changes. On the contrary, the modifications intensify the prevalence of the ‘new’ feature in the dialect sectors corresponding to the areas where Scandinavian presence was, eventually, felt most – the North and East Midlands. Clearly, the two sets favour SVO whichever scope of comparison is adopted.

The aim of my research was to test how confidently one could connect the morphological simplification to the subsequent syntactic reanalysis. Therefore, the data retrieval strategy kept the two processes in the same (syntactic) environment, namely, involving sequences of elements which were (originally) inflected. Whereas, the ready provided dialect information enabled to show that the feature distribution was confined to specific linguistic regions and that the contact situation in question was a crucial factor at play. Finally, the use of additional (statistical) measures such as the coefficient of variation (CV) (Frank & Althoen 1994: 58–59) not only turned out to be beneficial, by properly translating figures into facts but, at times, it even proved indispensable when the amount of raw data available was insufficient to draw valid conclusions.

Comparative Studies of Modern British and American English Constructions

The more data, the better:

The English comparative correlative construction in British and American English

Thomas Hoffmann (University of Regensburg)

One of the most interesting constructions of English is the comparative correlative (CC) construction (also known as the “covariational-conditional construction”, cf. e.g. Fillmore, Kay and O’Connor 1988; McCawley 1988; Culicover and Jackendoff 1999; Borsley 2004; Den Dikken 2005; Sag 2008):

- (1) [*the* []_{comparative phrase1} (clause)]_X [*the* []_{comparative phrase2} (clause)]_Y
a. [*the* [*more data*]_{comparative phrase1} *you have*]_X, [*the* [*better*]_{comparative phrase2}]_Y
b. [*the* [*more*]_{comparative phrase1} *she got to know him*]_X, [*the* [*less*]_{comparative phrase2} *she liked him*]_Y

Yet, while the construction’s syntax and semantics have received considerable attention in the literature, so far no in-depth usage-based analysis has been provided. Drawing on corpus data from British as well as American English (namely the BROWN, LOB, FROWN, FLOB set as well as the British component of the International Corpus of English project, ICE-GB), the present talk will try to close this gap.

A Hierarchical Configurational Analysis (HCFA; Gries 2008: 242-54) of the data reveals no significant effect of the factors VARIETY (British vs. American English) or YEAR (1960s vs. 1990s). However, with respect to the comparative phrase the HCFA shows that in both clause X and clause Y AdjPs are independently favoured (cf. (1), in which both comparative phrases are AdjPs), while NPs (such as in (2)) occur less frequently than expected by chance in either slot:

- (2) The [emptier]_{AdjP} our hearts become, the [greater]_{AdjP} will be our crimes. (BROWN--F42)
(3) The [more grass]_{NP} they can be persuaded to eat, the [more milk]_{NP} they give. (LOB--E37)

Furthermore, a covarying-collexeme analysis (using Gries’s (2004a) *Coll.analysis 3* script) yields several significant substantive collocations of comparative phrases across the two causes (including *the* [greater]...*the* [greater].../*the* [sooner]...*the* [better].../*the* [more]...*the* [more]...).

Finally, on top of phenomena such as the relative frequency of BE-deletion (*The longer the line is, the greater is the delay* (FLOB—J70)), subject-auxiliary inversion (*The more America declines, the better will Clinton's chances be to become President*. (FROWN--A14)) or the presence of a *that*-complementizer (*the more that is consumed the better* LOB--J42), we also discuss the status of double-bind structures (Franck 1985) such as (4) in which the second clause Y functions as the dependent variable (cf. Goldberg 2003: 220) of the first CC construction (*the more opaque that atmosphere is*]_X, [*the less conductive it is*]_Y) as well as the independent variable of the second CC constructions ([*the less conductive it is*]_Y [*the bigger the temperature difference you need to cross it*]_Z):

- (4) And [*the more opaque that atmosphere is*]_X, [*the less conductive it is*]_Y
[*the bigger the temperature difference you need to cross it*]_Z (ICE-GB:S2A-043)

As the sound files from the spoken component of the ICE-GB show, it is the CC constructions’ standard intonational contour (with a rise on the comparative phrase of the first clause, a fall on the comparative phrase of the second clause and falls at the end of both clauses) that allows for structures such as (4): for while Y is syntactically the dependent variable for the X-Y CC, its intonation clearly marks it as the independent variable for the following Z clause (since there is a rise on *the less conductive* and a fall on *the bigger*).

Comment clauses on the move: an account of current change

Gunther Kaltenböck (University of Vienna)

While the exact origin and historical development of comment clauses such as *I think*, *I suppose*, *I guess* are a matter of some discussion (cf. Thompson & Mulac 1991, Brinton 1996, 2008, Fischer 2007), it is generally agreed that they have grammaticalized from fully lexical clauses into epistemic markers (e.g. Brinton 2008, Traugott 1995, Van Bogaert 2009). The present paper explores to what extent there is evidence of further grammaticalization of comment clauses in recent decades and discusses methodological problems of tracing ongoing change. For comment clauses the difficulty lies in the scarcity of appropriate data from different periods of present-day English. It is possible, however, to compare results from various sources (such as the *Diachronic Corpus of Present-Day Spoken English* and the *Corpus of Historical American English*) and in this way draw conclusions about their current development.

The study investigates a range of different parameters, such as overall frequency and distribution (initial, medial, final), use of the *that*-complementizer, semantic-pragmatic scope over the host construction, collocation patterns, and pragmatic functions, which show that a comment clause such as *I think* is increasingly used not so much as an epistemic marker but as a textual/interactional device (cf. Kärkkäinen 2003, 2010, Kaltenböck 2010). This erosion of the epistemic function of *I think* can be linked to an increased use of variant forms in recent years, notably *I'm thinking*, *I just think*, *I'm guessing*, which seem to be recruited as epistemic markers to compensate for the fading modal meaning of *I think*.

To account for the development of comment clauses, the paper finally argues for a construction grammar approach which places comment clauses in a larger constructional network with taxonomic links to related constructions, viz. the matrix-object clause (transitive) schema and the sentence adverbial schema. This view can account for the advance of *I think* from an epistemic to a general pragmatic marker as well as the use and retention of the *that*-complementizer.

**Conceptual proximity and the positional
variation of directional modifiers in English**
Turo Vartiainen (University of Helsinki)

The ordering of modifiers in the noun phrase has received much attention both in typological and language-specific studies (see e.g. Bolinger 1967; Dixon 1982). In the studies of English, for example, the typical ordering of nominal (pre-)modifiers has been explained by semantic classes (e.g. Dixon 1982; Quirk et al. 1985), the discourse function of the modifiers (e.g. Hopper and Thompson 1984) or the modifiers' relative subjectivity (e.g. Breban 2006). The variation between premodifiers and postmodifiers in English, on the other hand, has not received much attention, which may be a result of the fact that postmodifiers were used only marginally already in the Middle English period (Raumolin-Brunberg 1994, Fischer 2006). In Present-Day English, the use of postmodifiers is generally seen as unprincipled, reflecting the "bits and pieces of those rules that were left over when decay set in" (Fischer 2006: 254; see also Huddleston & Pullum 2002: 445).

In this paper, I will focus on a sub-class of modifiers that may still occur both prenominally and postnominally in Present-Day English. I tentatively call these modifiers *directional modifiers*, viz. modifiers that are used to identify a discourse referent by pointing to its direction either spatially, temporally or textually. These modifiers include words like *following*, *coming*, *above*, *below*, *past*, *ensuing* and *foregoing* – words that formed a syntactically motivated group already in the ME period: contrary to the observed trend that favoured prenominal modifiers, these words almost exclusively occurred in the postmodifying position in Middle English, and came to be used in the prenominal position gradually over the Early Modern English period.

I would like to suggest that the variation between the prenominal and the postnominal uses can be explained by conceptual proximity. Specifically, I will suggest that the referents of the heads of prenominal modifiers are typically construed as proximal, which means that they are more clearly defined in the speaker's mind than their postnominal counterparts. As a result, prenominal directional modifiers tend to point to clearly-defined sets of referents, which are temporally, spatially or textually proximal, whereas postnominal modifiers usually point to more open-ended sets of referents, which are temporally, spatially or textually distal. This effect can be seen by comparing the following examples:

- (1a) During *the coming days* Americans will be throwing away some 35 million Christmas trees, creating an urban disposal nightmare. (a rather well-defined set of few days; BNC AAT 314)
- (1b) We will be watching for your reports in *the days coming*. (an open-ended set of days; COCA)
- (2a) In *the past days* people had begun cautiously to return to the moor, but not this evening and not here. (a rather well-defined set of few days, recent past; BNC FU2 848)
- (2b) The Sheikha again said that it was a story from *the days past* but Halema did not say another word. (an open-ended set of days, remote past; BNC CDX 906)

I will offer further support for the "proximity hypothesis" by presenting a diachronic corpus study of the most frequent directional modifiers, *following*. I will show that the change from the postnominal to the prenominal position occurred at different times in the textual and the temporal functions of *following* (e.g. *the following examples*, *the following Monday*) and that the order of change can be explained by conceptual proximity.

Contact or Comparisons of English and Related Languages

Interpreting modal utterances in English and Swedish

Anna Wärensby (Lund University)

The interpretation of modal verbs continues to pose both theoretical and descriptive problems. Many studies on modal verbs appeal to notions of context (e.g. Kratzer 1991, Coates 1983, Papafragou 2000). Few studies, however, undertake a task of actually defining these contexts. This study is usage-based. It investigates a large number of naturally occurring examples in order to outline patterns of interpretation of utterances containing modal verbs (see Author X 2006).

It is based on approximately 3,000 examples of the English *must* and *may*, and their Swedish counterparts *måste* and *kan*. These examples are extracted with co-text from the English-Swedish Parallel Corpus. The data are subjected to a data mining analysis using 21 different contextual variables, including time reference, type of subject, type of verb, aspectual modification.

The results of this analysis support the hypothesis that disambiguating utterances containing modal verbs is a process where certain variables or combinations of variables are of crucial importance, whereas others are not. 96.5% of examples containing *must* and 94.2% of *måste*-data are correctly predicted as being epistemic or deontic. 77.6% of *may*-data and 66.1% of *kan*-data are correctly predicted according to the different interpretations these utterances receive in context. Furthermore, the patterning of the contextual variables investigated with respect to modal interpretations is similar in the two languages under investigation, English and Swedish.

Some of the examples, however, have eluded data analysis. Most of these examples are indeterminate. Contrary to what is observed in the clearly epistemic and deontic examples, in the indeterminate examples there is no apparent systematicity in the co-occurrence patterns of the variables discussed.

The borrowability of English: A method for measuring the intrusion of English in European languages

Eline Zenner, Dirk Geeraerts, and Dirk Speelman (University of Leuven)

In this paper, we focus on the question which features influence **variation in the success and popularity of English loanwords**. Specifically, based on a case study involving the use of 125 English person reference nouns (e.g. *babysitter*, *manager*) in Dutch, we argue that a comprehensive and reliable study of the borrowability of anglicisms needs to overcome a number of methodological difficulties in existing anglicism research. In a broader context, we hope to show that it may be interesting to put anglicism research back on the agenda of English linguistics: as the English as a Lingua Franca paradigm has been growing more stable over the last decades, it has overshadowed anglicism research, leaving no room for the kind of methodological improvements that we will talk about.

More concretely, we argue that current studies are in need of the following four methodological improvements (all of which are transferable to other studies on borrowability, lexical variation and diffusion and lexical innovation). Firstly, we stress the importance of **sufficient data** (see Geeraerts 2010). So far, corpus-based accounts of anglicisms have mainly been based on limited samples, as researchers usually extract anglicisms manually from their corpora (e.g. Fink 1997). Our study, on the other hand, is based on a lemmatised and syntactically parsed newspaper corpus which contains over one billion words ranging from 1999 to 2005, and which represents the two national varieties of Dutch (Belgian Dutch and Netherlandic Dutch).

Secondly, we introduce a **new measure to compute the success of an anglicism** in a given corpus. The method (a derivative of the profile-based method; Speelman *et al.* 2003) is concept-based and is hence less susceptible to the influence of topic-specificity than the traditional type and token counts used in borrowability studies (see Van Hout & Muysken 1994 for other drawbacks). It defines the success of an anglicism as the corpus frequency of the anglicism, relative to the total frequency of the English loan word and its denotational synonyms (see example in Table 1).

Synonymous expressions for the concept <i>hooligan</i>	Token frequencies in the Dutch corpus
<i>hooligan</i>	9337
<i>voetbalvandaal</i>	611
success-rate for <i>hooligan</i> : $9337 / (9337 + 611) = 93.9 \%$	

Table 1

We semi-automatically calculate these success-rates for approximately 125 English person reference nouns across eight subsections of the corpus, giving us 1000 data points based on over 30 million tokens. The identification of synonyms is based on lexicographical sources and on automatic detection by means of word space models (see Lin 1998 and Peirsman *et al.* 2007)

Thirdly, we wish to **expand the number and nature of the explanatory variables**. Where traditional anglicism research has been preoccupied with the influence of structural features (like word class) on borrowability, we broaden the spectrum: we verify the effect of conceptual parameters (lexical field, concept frequency, conceptual newness), extra-linguistic parameters (regional and register variation) and several additional parameters (age of the loanword, spelling adaptation ...). Special attention is paid to defining objective operationalizations of these parameters.

The final and most important improvement is the application of **multivariate statistical techniques**. Using a logarithmic transformation of the success-rate, we perform mixed-effect regression analyses to find out which of our explanatory variables have an actual effect on the success of anglicisms when taking the combined effect of all variables into account. Results show the overwhelming importance of the conceptual features, leaving little to account for by extra-linguistic differences. The effect of the more structural features is negligible.

Psychological Aspects of English Syntax

A cognitive approach to obligatory subject-dependent XVS constructions in English Carlos Prado-Alonso (University of Santiago de Compostela)

Over the past decades, *subject-dependent* inversion —constructions in which the subject follows the entire verb phrase in a declarative clause, as in “On the near corner was *Herb’s Gas Station*” or “Upstairs was *a bedroom and a bathroom*”— has been the subject of extensive research from a generative (cf. Bresnan and Kanerva 1992; Schatcher 1992, among others), a functional (cf. Birner 1996; Dorgeloh 1997; Kreyer 2006; Prado-Alonso 2009), and a cognitive linguistic perspective (cf. Fillmore 1999, Chen 2003).

Most studies on English full inverted constructions base their classifications on the preliminary distinction between two main types of subject-dependent inversion: *non-obligatory* subject-dependent inversion and *formulaic* or *obligatory* subject-dependent inversion. Non-obligatory subject-dependent inversions are constructions in which the addressor can opt for either the inverted word-order or its canonical counterpart (cf. 1). By contrast, obligatory subject-dependent inversions are those constructions which are obligatorily triggered by certain fixed preverbal constituents, and are not replaceable in context by a comparable clause with canonical word-order because this is grammatically unavailable or conveys a different meaning, as illustrated in (2)-(3) respectively.

- (1) a. On his right was the mountain.
b. The mountain was on his right.
- (2) a. Nor are the monumental display scripts of the two books particularly close.
b. * The monumental display scripts of the two books are nor particularly close.
- (3) a. Here is our correspondent Mark Tully.
b. Our correspondent Mark Tully is here.

The analysis of obligatory or formulaic subject-dependent inversion has been neglected in the literature, with most studies focussing only on the non-obligatory type (cf. Kreyer 2006, among others), since it has been argued (cf. Green 1982; Chen 2003) that there seems to be no linguistic variation involved in the use of obligatory subject-dependent inversion. Here, I defend the view that obligatory subject-dependent inversion is a productive, highly structured construction which is worthy of serious linguistic investigation. I will argue that the different obligatory subject-dependent inversion types distinguished in the literature are clear instances of *constructions* as understood in the *construction grammar* framework. Through an in-depth analysis of four computerised written and spoken corpora – namely FLOB, FROWN (cf. Hofland et al 1999), CSPAE (cf. Barlow 2000), ICE-GB (cf. Nelson 1988) – I will show that despite their formal and functional dissimilarities, these inversions still relate to one another in systematic and predictable ways. I will further show that in fact the inversions are grouped in relation to a unit in the schematic network which is naturally most salient – the prototype (cf. Langacker 1987: 492) – and form with it a family of nodes – extensions from the prototype (cf. Langacker 1991: 548) – in the system. In sum, the analysis here will show that obligatory inversions are *constructions* which form an interconnected, structured system or network and are best understood with reference to different forms of *inheritance* (cf. Goldberg 1995: 72 ff).

**Learning Verb-Argument Constructions:
New Perspectives from Corpus and Psycholinguistic Analyses**
Ute Römer, Matthew O'Donnell, and Nick Ellis (University of Michigan)

Our experience of language allows us to converge upon similar interpretations of novel utterances like “the ball *mandools* across the ground” and “the teacher *spugged* the boy the book.” You know that *mandool* is a verb of motion and have some idea of how *mandooling* works – its action semantics. You know that *spugging* involves some sort of transfer, that the teacher is the donor, the boy the recipient, and that the book is the transferred object. How is this possible, given that you have never heard these verbs before? Each word in these utterances contributes individual meaning, and the verb meanings in these Verb-Argument Constructions (VACs) are usually central. But the larger configuration of words (the construction) has come to carry meaning as a whole (Goldberg, 2003, 2006). The VAC as a category has inherited its schematic meaning from all of the examples you have heard. *Mandool* inherits its interpretation from the echoes of the verbs that occupy this VAC – words like *come*, *walk*, *move*, ..., *scud*, *skitter* and *flit*. Our research explores these processes of usage-based VAC acquisition from large- scale corpus analysis and data from psycholinguistic experiments.

In the corpus analyses, we are constructing an inventory from searches of a dependency-parsed version of the British National Corpus (BNC) for specific VACs previously identified in the *COBUILD Grammar Patterns* volume (Francis, Hunston, & Manning, 1996). Our method applies the following steps. For each VAC, such as the pattern V(erb) *across* n(oun phrase), we generate (1) a list of verb types that occupy the construction (e.g. *walk*, *move*, *skitter*). We tally the frequencies of these verbs to produce (2) a frequency ranked type-token profile for these verbs, and we determine the degree to which this is Zipfian (e.g. *come* 474 ... *spread* 146 ... *throw* 17 ... *stagger* 5) (Zipf, 1949). Because some verbs are faithful to one construction while others are more promiscuous, we produce (3) a contingency-weighted list which reflects their statistical association (e.g. *scud*, *skitter*, *sprawl*, *flit* have the strongest association with V *across* n). We use (4) WordNet disambiguated senses of these verbs in the sentences in which they occur to analyze (5) measures of the degree to which there is semantic cohesion of the verbs occupying each construction (e.g., semantic fields TRAVEL and MOVE most frequent for V *across* n). We measure the degree to which each VAC is more coherent than expected by chance in terms of the association of their grammatical form and semantics, and assess the consequences of this for their acquisition in terms of what is known of the psychology of learning.

In the psycholinguistic experiments we test these predictions and measure the verb-to- construction knowledge of both native speakers of American English and non-native speakers. We use free association tasks to direct subjects to think of the first word that comes to mind to fill the V slot in a particular VAC frame. The range of the verbs that they generate, and their speed of access, inform us about the representation of these VACs in the human mind. We show how the statistical factors in our inventory predict this. Our work illustrates the productive synergy of corpus linguistic and psycholinguistic methods and findings.

Aspects of seriality in language. Or: On the order of schemata in our brain

Rainer Schulze (Leibniz University of Hanover)

Starting from the observation that meaning does not reside in individual words but in the phrase or construction, this paper will focus on the examination of recurring phrases or constructions in language and on adverbial *considering* and *to somebody's chagrin* in particular, as shown in the following examples:

- (1) He thanked Mitch again for the tattoo and said it was lasting pretty well, considering. (BNC: 2604/C86/W_fict_prose)
- (2) Yes, the father showed up to identify. Very composed, considering. (BNC: 2770/H8L/W_fict_prose)
- (3) Atlanta Georgia is to host the 1996 Olympics much to the chagrin of Athens and Manchester. (BNC: KRT/S_broadcast_news)

Despite a number of stimulating studies focussing on emotive language and stance adverbials in particular (Bednarek 2008, Biber/Finegan 1989 or White 2002), the role of distributional and environmental factors in recurring patterns, as made evident in Martin's systemic-functional appraisal theory, Biber's theories concerning register variation, Hunston's local grammar approach or Hoey's lexical priming assumptions, need to be incorporated in a construction grammar framework and will be explored in some greater detail. Given that all the components of a construction "...incorporate various contextual clues" (Fried 2009: 88) or that "...everything that happens to the meaning of a gram happens because of the contexts in which it is used" (Bybee/Perkins/Pagliuca 1994: 297), I will discuss some of the challenges that the inclusion of distributional and environmental factors within the construction grammar framework presents. Based on previous work in cognitive psychology, cognitive linguistics-*cum*-construction grammar, lexicogrammar, pattern grammar or lexical priming, I will argue that the non-linguistic notion of schema can successfully be reconciled with recurring patterns or constructions in language. The results of the study will show that the corpus-based analysis of selected evaluative markers such as *considering* or *to somebody's chagrin* offers several advantages. Firstly, by exploring the surrounding of a pattern or functional structure (Schulze forthcoming) in a limited amount of data, we are not only able to find evidence for the specific readings of an evaluative marker, but we are also able to enrich the description of the evaluative marker with information from a series of different dimensions and add a pragmatic viewpoint by revealing both an explicit and implicit reading of the marker. In doing so, we will postulate the principle of seriality, i.e. the order in which some recurrent phrases and constructions follow each other. This order is determined by their semantic and pragmatic affinity of the phrases and constructions concerned, which is, in turn, determined by the kind of reality they refer to. This linguistic affinity has direct mental consequences because the pertinent phrases and constructions are obviously remembered as parts of schematic information, not by themselves. Instead of the single phrase or construction, the series of phrases and constructions is obviously made the unit of mental representation. The second advantage offered by the corpus-based approach lies in the possibility of quantifying results, i.e. comparing the frequency, productivity and creative use of selected evaluative markers, which is especially interesting in view of British and American English contrasts. This assumption will be supplemented by a number of research questions: Are *considering/to somebody's chagrin*-constructions entrenched despite their low frequency of occurrence, and simply activated? Is the preceding and following context structured in such a way that we can assume that the *considering/to somebody's chagrin*-constructions are anticipated and planned well ahead of their actual occurrences? Thirdly, my intention will be to show that the combination of functional approaches and corpus-linguistic research work will result in a promotion of usage-based studies within the construction grammar framework.

**“You’ll catch your death of cold”:
The Influence of Lexical Conflation on Beliefs About Causation**
Laurel Smith Stvan (University of Texas, Arlington)

A prevalent trope in vernacular health discussions is a connection between people being cold and their catching a cold, presented as a claim that the first state causes the second. At the same time, the number of works written by journalists and public health writers to debunk this connection further underscores how widespread the idea is (Noah 2007, Wanjek 2003, Zuger 2003, *inter alia*). The current paper refocuses this discussion in the following two ways: First, using corpus data I present a quantitative baseline showing how prevalent this connection is within health discussions in English. Secondly, I posit that *cold* is only one of a wide set of words for which “lexical conflation” of polysemous terms presents an account for the intertwining and over literalization of the concepts involved (Stvan 2007, 2008).

Two data sources provide material for this analysis: the first is a general corpus of English, COCA (the Corpus of Contemporary American English). In addition, a more genre-focused dataset was pulled from CADOH (the Corpus of American Discourses on Health). Results--in line with the same causation argued against by Benjamin Franklin 250 years ago (Green 2008)--show common themes for Starting State Predicates (*cold, chilled, wet, drafty*) and Endstates (*a cold, your death, sick, pneumonia*). These are illustrated by the corpus tokens below:

- 1) I was rescued. The river was full of **ice**, but I didn't even **catch a cold**.
- 2) One of the worst things she used to do was turn off the **air conditioner** on 95 degree days because she didn't want to **catch a cold**.
- 3) You could go outdoors with **wet hair** every day and you wouldn't **catch a cold** -- provided you're a hermit,
- 4) “Just look at you! **Bare feet!** You'll **catch your death of cold** is what you'll do!”

Not all mentions of “catching a cold” talk about its cause. And conversely, not all descriptions of being cold or wet lead to discussions of catching a cold. However, of the triggers mentioned for catching a cold, the majority fell into the cold/wet category. The breakdown is as follows: Of the 233 “catch a cold” tokens in the dataset, 19.7% mention no cause; 25% mention the cause as viral contagion; 51% mention being cold, wet, or chilled as the cause; and 3.8% of them note some other cause.

Instantiations of this belief that *cold*ADJ leads to *cold*NOUN for English speakers are explored linguistically through interactions of iconicity, metonymy, and core meanings in polysemy (Williams 1992), while no case is found for underlying conceptual metaphors involving temperature (Lakoff 1987, Lakoff & Johnson 1980). Furthermore, this same conflation is shown to occur for ‘cold’ words from separate language families.

The idea that *cold* leads to *cold* has a shared etymology, but a false etiology. While beliefs about catching a cold may have originally brought about the extended sense (e.g. the noun *cold*), the similarity of the current terms is posited as an influence for much of the continued belief about the connections between the concepts.

Perceptual Dialectology

A new method for dialect recognition and rating in perceptual dialectology

Chris Montgomery (Sheffield Hallam University)

This paper will deal with a new method for collecting and processing data relating to the placement and ratings of voice samples by non-linguist respondents as part of a complete methodology for perceptual dialectology study. The method is designed to respond to issues present within Preston's (1999, xxxiv) 'principal techniques' for the study of perceptual dialectology.

Preston (1999, xxxiv) states that 'dialect recognition' tasks and 'draw-a-map' tasks form part of the 'principal techniques' for perceptual dialectology investigation. However, in discussions and demonstrations of the implementation of 'dialect recognition' tasks (Niedzielski and Preston 2003; Preston 2010), it is clear that these tasks diverge significantly from other 'principal techniques'. The 'dialect recognition' tasks undertaken by Preston restrict respondents by requesting that they make placements for voice samples only along a north-south continuum (Niedzielski and Preston 2003). Therefore, in their approach, 'recognition tasks' as used by Preston do not afford respondents the 'free choice' that the 'draw-a-map' element of the 'principal techniques' does.

'Draw-a-map' tasks are 'free-choice' tasks which ask respondents to draw lines on blank or minimally detailed maps indicating where they believe dialect areas to exist (Preston 1999, xxxiv). They are well known and much repeated in perceptual dialectology (Long 1999a; Fought 2002; Montgomery 2006; Bucholtz et al. 2007). The 'free choice' nature of the 'draw-a-map' task is a strength of the method, although it does present problems in terms of analysis, some of which have been addressed through computerisation of such data by Long (1999b), and Montgomery (2010). Such computerisation results in composite maps which indicate 'perceptual isoglosses' indicating the majority of informants' perceptions of the placement and extent of dialect areas.

This paper will discuss a free choice 'dialect recognition' task which is designed to better complement the 'draw-a-map' task. The task is carried out by respondents in an online environment. The new technique allows respondents to listen to voice samples and then to rate and place the samples on a map with a mouse click. The ratings and placement data are collected along with biographical data entered by each respondent before commencing the task. Data are aggregated automatically and generalised maps are produced which indicate through the use of 'hot spots' where each voice sample was placed by a number of respondents. These 'hot spot' maps can then be compared with composite draw-a-map perceptual isoglosses.

The data discussed in this paper were provided by respondents from the British mainland over a six-month period. The respondents were reacting to six voice samples taken from locations in the British mainland. Comparisons between 'hot spot' maps and perceptual isogloss maps generated from previous studies in mainland Britain (Montgomery 2006; 2010) will be discussed, and the extent of agreement between hot spot maps and perceptual isogloss maps investigated. It will be demonstrated that this new method for gathering ratings and placement data from respondents has much to offer the field of perceptual dialectology, not least in ensuring more completeness when investigating non linguists' 'mental maps of regional speech areas' (Preston 2002, 51).

Variationism

That's not just counting: Why empirical studies of prescriptive rules should be variationist Don Chapman (Brigham Young University)

This paper responds directly to the theme of “Methods Past and Current” by examining variationist and corpus methods for evaluating the relative importance of individual prescriptive rules. There is nothing new in examining a prescriptive rule against a corpus. The notion that correctness should be measured by actual use has been around for centuries, and with the development of large-scale corpora, the means of examining the actual language of speakers and writers has become feasible (Finegan 2001). Nor is it a new idea that variation is what gives rise to nearly all prescriptive rules: in a typical prescriptive rule, one variant is promoted (like *different from*) and all the other variants are proscribed (like *different than* and *different to*). The new (or at least insufficiently acknowledged) idea presented in this paper is that just counting up occurrences of prescribed vs. proscribed forms is not enough for an empirical examination of prescriptive rules; we need to do that counting within traditional sociolinguistic variationist methods.

There are two main points that argue for a variationist approach. The first is that the prescriptive tradition often makes potentially empirical claims that depend on comparison. When a usage handbook claims that “Standard American English is the dialect generally accepted by educated speakers and writers” (Ebbitt and Ebbitt 1990: x), it implies that some other variety is accepted by less educated speakers and writers. In this claim, different forms in English are implied to depend on different levels of education. The second point is that the distributions of variants in prescriptive rules will practically never be binary. If they were, they would not be noticed enough to prompt a prescription (cf. Finegan 1980). It is not as if one variety of English will always use the prescribed form and another variety will never use it; instead the two varieties will differ, if they differ at all, in the frequency of the prescribed form. This is precisely the type of data that typical sociolinguistic variationist approaches treat.

In a variationist approach, the prescriptions will be dependent variables, and the independent variables will be those factors that are posited in the prescriptive tradition to make a difference. In the example above, education would be the independent variable. Some other factors that are hinted at in the prescriptive tradition as making a difference are mode (speech vs. writing), register (business or news writing vs. academic writing), and style (e.g. careful, formal, unmonitored).

Implementing a variationist approach will be easier for some of these independent variables than others. Large-scale, general-purpose corpora, like the BNC and COCA, allow for comparisons between speech and writing, and across various registers. Implementing a variationist approach with education as a social variable will be much harder, since it would require a large corpus stratified by education level. Such a corpus has been hard to come by, since most corpora depend largely on published sources, which are usually written and edited by those with at least some university education. But without a corpus stratified by education, the connection between education and the use of prescribed forms cannot be adequately empirically tested. Since prescriptive rules are so widely thought to correspond to education, this lack of a corpus is particularly regrettable. The next step in subjecting prescriptive traditions to full empirical scrutiny, then, will come with corpora that can allow a variationist approach with education as a social variable.

Morphological Methodology for a Rapidly Reconfigured Variable: *Be Like* in Appalachia

Kirk Hazen (West Virginia University)

With the study of morphological variables, the methods of investigation must explicitly take into account the data under study. Buchstaller (2009) describes the contentious practice of extending methodologies used for phonological analysis to morphological variables, specifically referencing the debate about the construction of *the variable* as an entity. This debate is directly addressed through examinations of quotative variation in one of the most renowned US vernacular dialects. At the end of the twentieth century, a rapid change in the quotative system unfolded across the English speaking world. Specifically, quotative *be like* rose to be the primary choice for many younger speakers in urban and suburban areas. Buchstaller and D'Arcy (2009) argue for the weak transfer of global constraints to fit the specific sociolinguistic profiles and previously entrenched quotative systems of local communities. Importantly, for a morphosyntactic variable such as quotatives, two questions appear: 1) can the community-based social groupings of traditional variationist analysis handle the inconsistent production amongst speakers? 2) Does the traditional concept of the *variable* allow for the best analysis of morphological variation? To achieve optimum analysis in this paper, methodological decisions were made about the nature of the variable and the delineation of age groups.

With an eye to these methodological decisions, this paper examines the social and linguistic constraints on the ascendance of quotative *be like* in Appalachia to track the pathways of language variation and change. This paper further demonstrates how external connections and stereotypes of *be like* guide both its rise and the suppressed rates of other quotative verbs. An evenly balanced corpus of sociolinguistic interviews from 67 native Appalachians was meticulously examined for seven dependent variants, 14 linguistic factor groups, and six social factor groups. A logistic regression analysis of 2,666 tokens of quotative verbs yields a clear view of the statistically significant language variation patterns for these speakers. Unfortunately, the distribution of quotatives is not even across the subject pool, with some speakers producing 200 quotatives in an interview and some producing less than 5. The most decisive trend is the ascension of *be like*: In general, older speakers produce less than 1% *be like*, in sharp contrast to the middle-aged group (31%) and the youngest speakers (69%). To some extent, for these Appalachian speakers, the rise of *be like* came at the expense of *say*, which declined from 70% to 13%. This reversal of quotative verbs did not occur in a smooth curve. On closer inspection, the females and males took different paths. Crucially in the middle-aged group, the first to adopt the new, external variant *be like*, the females dropped their rate of *say* 50% in order to adopt *be like*, whereas the males differentiated themselves by nearly doubling their rate of zero quotatives and adopting *be like* only 15% of the time. In the end, it appears that Appalachians are aligning themselves with US norms, despite the vernacular tradition of the Appalachian region. To methodologically handle this rapid change, given the unequal production of quotatives, the normal variationist method of adopting set age groups had to be dismantled and a continuous factor group of age was implemented. Additionally, Buchstaller's (2009) suggested method of "closing the set" allowed the analysis to handle four morphological variants of the dependent variable. With this reconfiguration, the analysis reveals that the path for quotative language variation and change was guided by linguistic constraints but divided sharply by social differences.

Variation and Change in the north-west of England

Sandra Jansen (University of Duisburg-Essen)

GOOSE-fronting has been described for many varieties of English around the world and especially for those in the U.S. (e.g. Hall-Lew 2009, Baranowski 2006, Fridland 2008, Fought 1999; see Mesthrie (2010) on the situation in South African English). It has been reported for various areas in England (e.g. Britain 2008, Harrington et al. 2008, Torgersen 1997). Docherty (2010: 67) states in an overview on recent changes in British English:

Perhaps the most striking of these [vowel changes] is the fronting of the GOOSE and GOAT vowels by younger generations of speakers. Putting to one side those varieties where GOOSE is already fused with FOOT [...], there are widespread reports of moderate to substantial fronting of GOOSE together with the production of much less marked lip-rounding/protrusion [...].

Similarly, Ferragne and Pellegrino (2010), who analysed 13 accents across Britain, suggest a general fronting of GOOSE but also comment on the relationship between GOOSE and FOOT in the various varieties.

Carlisle, a city of 101,000 in the north-west of England, provides an interesting setting for linguistic study. The proximity to the Scottish border and the lack of neighbouring urban conurbations make it a unique location in the British Isles. Yet, only very recently researchers have started to devote their attention to this area (Montgomery 2006, Llamas, Watt et al. 2008-2011, Jansen in preparation).

The aim of this talk is to present results on the age, gender and social variation of realisations of the GOOSE lexical set based on data by 42 speakers of Carlisle English. Acoustic phonetics is used as a means of analysis as it “enables us to capture relatively fine-grained differences in vowel quality across different speakers and contexts” (Hay and Drager 2007: 92). Furthermore, data for GOAT and FOOT is taken into account in order to identify fronting tendencies in these two vowel sets.

Pragmatics

Variationist sociolinguistics meets variational pragmatics: Investigating responses to thanks

Markus Bieswanger (University of Flensburg)

This paper addresses a core methodological issue in the emerging discipline of variational pragmatics, which “is conceptualized as the interface of pragmatics with variational linguistics” (Schneider & Barron 2008:1). Despite the known drawbacks of many kinds of experimental data – particularly the use of discourse completion tasks (DCTs) for data collection (cf. Barron 2003:84-86, Blitvich 2006:216-218) – much research in pragmatics still relies on such data for reasons of efficiency and convenience (cf. list of contributions in Schneider & Barron 2008:23). The collection of natural data, especially with controlled sociolinguistic variables, is considerably more challenging and time-consuming. With respect to variational pragmatics, I will compare the use of DCT-data to the use of electronic corpora, such as the *International Corpus of English (ICE)*, and natural data collected with the help of Labovian-style methodology common in variationist sociolinguistics. I will then present the results of an empirical study and explain why I am convinced that the latter method of data collection gives us the best insight into actual authentic language use.

Research in pragmatics has repeatedly addressed responses to thanks as a speech act that fulfils an important social function (cf., e.g., Aijmer 1996; Schneider 2005). Wolfram & Schilling-Estes (2006:93) claim that the way “[...] speech acts are carried out and the conditions under which they are considered to be appropriate varies considerably across cultural groups.” For the study underlying this talk, natural language data was collected in identical situations in several cities in different countries where English is the native language of the majority of the population. Each data set consists of a gender-balanced corpus of 60 responses to thanks, controlling other sociolinguistic variables such as age, ethnic group and socioeconomic status as precisely as possible during the data collection process. The results of the analysis of the natural data will be presented and compared to Schneider's (2005) DCT-based study of responses to thanks, identifying a number of fundamental differences.

Noun ellipsis in context:
Pragmatic factors determining variation in the realization of head nouns
Christine Günther (University of Osnabrück)

According to Quirk et al. (1985) noun phrases in English “consist of a head, which is typically a noun, and of elements which (either obligatorily or optionally) determine the head and (optionally) modify the head...” (ibid: 62). Hence, the typical head of an NP is a lexical noun, as illustrated below.

(1) the green **shirt**

The head noun *shirt* denotes a kind whereas the NP refers to one particular instance of the type. If, in ongoing discourse, reference is made to further types of that kind, three options arise: a) the lexical noun can be repeated, b) the head can be replaced by *one*, and c) the head can be empty.

- (2) a. the green shirt and the black **shirt**
b. the green shirt and the black **one**
c. the green shirt and the black [e]

The latter option, i.e. the use of an empty nominal element, also referred to as “noun ellipsis”, is highly restricted in English when it comes to adjectival modifiers. More formal approaches even speak of “quasi-absence of noun ellipsis in English” with modifying adjectives (Sleeman 1996: 14), claiming that the construction determiner + adjective must be completed by one. However, it is also claimed that this strict requirement is relaxed in strongly contrastive contexts, as witnessed in the oft-quoted example from Halliday & Hasan (1976): I like strong tea. I suppose that weak is better for you (cf. Bouchard 2002: 225).

This gives rise to the following questions: first, in which contexts is an empty noun licit, and, second, which factors govern the choice of one of three ways to realize a nominal head? In order to explore these issues a study of the British National Corpus was conducted. The aim of this paper is two-fold: answers to the above questions will be sought while at the same time the applicability of corpus studies to the analysis of empty categories will be discussed.

In this paper the choice of nominal head in two different contexts will be presented: contexts featuring the antonymic adjective pair old-new (cf. example (3)) and coordinated NPs as illustrated in (4):

- (3) a. ... a *new* bridge was completed beside the remains of the *old* [e], ... (B0A 865)
b. ... some can remember *old* faces but cannot learn *new ones*. (A0T 1201)
c. All this *new* history makes *old history* even more important. (A2W 125)
- (4) a. As far as I was concerned there were two types of women, *the true type* and *the failed type*.
b. There are two views of this task, *the real one* and *the naive one*. (FNR 266)
c. This strategy can be divided into two approaches, *the forceful [e]* and *the logical [e]*. (CEF 705)

It will be argued that the notion of “contrastive context” is not fine-grained enough to account for the licensing and the use of noun ellipsis. Rather, a number of factors have to be taken into consideration, such as the contextual limitation of a reference set, accessibility of the antecedent, the type of the antecedent, stylistic preferences and aspects of language economy. The interaction of these factors will be discussed for each type of head in order to shed light on the pragmatics of noun ellipsis and on variation in the use of head nouns.

You wanna check this out!
On the pragmatics of obligative *want to* – a wolf in sheep’s clothing?
 Meike Pfaff (University of Osnabrück)

In order to perform directive speech acts, speakers of English have a set of expressions to choose from, as there are for instance the use of the imperative (1), periphrastic expressions involving the modals *must* and *shall* (2), as well as the semi-modals *ought to*, *have to* and *need to* – each of them expressing different degrees of obligation as well as politeness. Another expression belonging to the latter, i.e. semi-modal group is the *want to* construction as found in (4) which is typically regarded a politeness strategy employed by the speaker to elegantly lay obligation on the addressee.

- (1) At the next traffic light take a left onto Boylston St!
- (2) When you come to Charles Street south, you must take a right turn.
- (3) The MIT is to our left – you have/ought/need to make a u-turn!
- (4) You want to take the Ted Williams Tunnel to get to Boston.

The aim of this paper is three-fold: drawing on corpus data from the BNC and the COCA, it discusses in a first step the modern use of the *want to* construction and its status along the lexical-to-modal continuum. In a second step, the paper seeks to examine the motivation behind the use of obligative *want to*, which so far has been treated as a rather recent phenomenon of the English language (see Krug, 2000; Mair&Leech 2006). To this end, it tackles the issue from two angles: coming from an onomasiological perspective, it discusses what may have motivated speakers to choose for the expression of obligation a verb of volition, especially as obligation and volition seem to represent but two antonymic concepts. In other words, is the *want to* construction the result of pragmatic inferencing as postulated by Krug (2000), who argues that “what somebody has to do in his or her own interest is what he or she is willing to do”? To this end, parallels from German will be discussed, which in a similar way utilizes a verb of volition in an imperative-obligative context:

- (5) German
Der Chef sagt, Du möchtest sofort in sein Büro kommen.
 the boss says you want at once in his office come
 ‘The boss wants you to come to his office at once.’

Taking on an etymological-semasiological perspective, the paper will trace the historical development of the lexical verb *want* so as to propose another motivation for its apparently ‘new’ obligative meaning. Based on a corpus study using the Corpus of Historical American English (COHA) as well as other resources it will be shown that obligative *want to* might instead be the manifestation of an older, pre-volitional meaning of *want* ‘be lacking’, dating back to the Middle English period.

- (6) Middle English (Ancrene Wisse, c. 1230)
a word ne schal þer wontin.
 a word NEG shall there lack.PROG.INF
 ‘not one word shall be lacking there’

Finally, this paper also brings up methodological issues and the question of the limitations of corpus research, especially when dealing with synonymy.

On how to integrate context into grammar

Alexander Bergs (University of Osnabrück)

Construction Grammar (CxG) defines constructions as holistic or very frequent form-meaning pairings at all levels of linguistic structure. The constructions of a given language form a structured inventory. Knowledge of language means knowledge of constructions and their network. Most CxGs assume that pragmatic information, e.g. co(n)textual information, can be integrated into constructions. However, only few studies have done this so far. This paper is concerned with how to put this theoretical claim into practice, i.e. the methodology of how to systematically integrate co- and contextual information into construction grammar. The exemplary test case will be some constructions in English signaling futurity. The four major ways of expressing futurity in present-day English are:

- (1-a) John WILL make a toast.
- (1-b) John IS GOING TO make a toast.
- (1-c) John MAKES a toast tonight.
- (1-d) John IS MAKING a toast (tonight).

All four constructions share a core of meaning (S to the left of E and R), but there are also complex co-occurrence restrictions and subtle differences in meaning: (1-a) makes general predictions about future events, (1-b) implies immediacy and/or plans underway; (1-c) implies “future facts”, (1-d) certain arrangements underway, though the future event could still be cancelled. In terms of co-occurrence restrictions, we find that (1-a) and (1-b) are both modifiable by time adverbials. Adding tomorrow makes (1-a) more certain, and it lowers the immediacy overtone in (1-b). (1-c) is almost unacceptable without modification by a time adverbial; (1-d) is ambiguous unless modified. Moreover, (1-a) basically allows for almost all subject types, animate (agentive) and inanimate. Still, first person subjects in this context can lead to ambiguities between epistemic readings (prediction), and root readings (volition, i.e. resolve and willingness). This effect is further strengthened if dynamic or intentional action verbs are present.

This paper explores in how what way these phenomena can be captured in CxG. It claims that the meaning differences are based not on the polysemy of single verbs or adverbials, as is often assumed, but on the construction AS A WHOLE and its context (e.g., shared (world) knowledge, co-text etc.). The ambiguities in (1-a) with first person subjects can be modeled as polysemy links between related constructions, since these constructions (epistemic and root) share the same syntactic configuration, but show distinctive meanings. These can be distinguished through contextual factors, or grammatical configurations: Ambiguity usually does not arise with other person subjects; these structures are uniquely ascribed to the epistemic construction. (1-d) only gains its specific meaning either in context or through the appropriate adverbials. In (1-c), the adverbial slot is obligatory in most cases, though, even here, cotextual factors might suffice to disambiguate the simple present construction from the future construction. However, on a critical note we also need to ask in how far we need to refer to grammatical factors, and when simple ad-hoc pragmatic reasoning suffices for the speaker/hearer in en-/decoding the meaning of the utterance.

It is thus suggested that future constructions have an (optional) syntactic and semantico-pragmatic slot which can be filled by either implicature through context or specific morphosyntactic items that provide that context. With reference to frame semantics and scripts, this paper develops a unified account for future constructions with and without (semi)auxiliaries in which different semantico-pragmatic functions are distinguished as the meanings of different but related construction types.

African and related diasporic Englishes

Gauging Variety Status in Diasporic Dialect Mixing: Toronto Jamaican as a Potential Source for a Canadian Black English

Lars Hinrichs (University of Texas, Austin)

Urban diasporic communities provide a site for the observation of dialect contact from initial stages of mixing, marked by frequent codeswitching (for the Jamaican community in Britain see e.g. Sutcliffe 1984; Sebba 1993) to the emergence of stable, fused lects by the third generation (Cheshire et al. 2008; Kerswill, Torgersen, & Fox 2008; note, however, that Jamaican Creole is only one of several diasporic tributaries to this new dialect). In the first linguistic approach to a diasporic community, the most pressing concern is to determine the dialectological status of the result of dialect contact. Specifically, one may wonder if the term 'variety' can be legitimately used in describing the mixture at hand. This paper reports on an ongoing analysis of sociolinguistic recordings from the Jamaican community in Toronto. The major tributary varieties to Toronto Jamaican Speech (ToJS) are Canadian English, Jamaican English, and Jamaican Creole.

This paper discusses to what extent we can speak of ToJS as a variety (yet). In their function as sociolinguistic indexes of racial group membership—as impressionistic macro-level descriptions have described them—Jamaican English/Creole speech forms have a powerful rival in African-American Vernacular English. Also, my interviews among members of the Toronto Jamaican community have shown that there are great differences among ideolectal patterns of combining the major input varieties of ToJS (Hinrichs, under review). At the same time, however, some speakers in my sample do show signs of speaking a stabilized, Jamaican-Canadian "fudged lect" (defined by Chambers & Trudgill 1998 as a stable lect combining two input lects through compromise phonetic forms; cp. Auer's definition of "fused lects," which focuses on sedimentation of mixed lects into lexical and grammatical stability, 1999). Also, youth language and the language of Canadian hip hop are sites in which stable combinations of Jamaican and Canadian phonetic features can be encountered, with conversational switching between Creole and English turning into a marginal phenomenon.

While ToJS has not yet reached the level of enregisterment (Agha 2007) and structural stability that the multiethnic inner-city dialect in London, described by Kerswill and colleagues (see also Ashton 2006), has acquired, I argue that in certain contexts, defined by criteria of high network density and positive language attitudes, Jamaican-Canadian fudged lects do arise that may in time give rise to a new Jamaican Canadian dialect.

This project is a case study in initial approaches to urban diasporic varieties. I argue that a methodological envelope combining the strengths of both structural/quantitative linguistics (through sociophonetic analysis) and of a qualitative, cultural studies-based perspective is the most promising in addressing this task.

**New ways of analysing the history of varieties of English:
Early Highlife recordings from Ghana**
Magnus Huber and Sebastian Schmidt (University of Giessen)

The growing interest in World Englishes has been accompanied by an explosive number of synchronic studies of these varieties on all levels of linguistic description. While the history of standard and, to a certain degree, non-standard varieties of Inner Circle Englishes (in Kachru's 1986 terms) has also received linguistic attention, diachronic investigations of Outer Circle (or post-colonial) varieties are still the exception. Very often, descriptions of the history of post-colonial Englishes are restricted to sociohistorical outlines from a macro-sociolinguistic perspective, with little if any reference to the linguistic structure of earlier stages of the varieties. One main reason for this lack of diachronic studies is the dearth of authentic historical data, either because there simply are none or because the sources have not been located and tapped yet. With regard to written sources, the situation is maybe less problematic since early travel accounts, diaries or memoirs of missionaries, traders and administrators often contain quotes from the variety, and at times there are even documents produced by speakers of colonial Englishes themselves (for West Africa, cf. the diary of Antera Duke, a late 18th century Nigerian slave trader; Behrendt et al. 2010). With due caution, such material can be used to give us an insight into the morphology, syntax and the lexicon of earlier stages of overseas varieties of English (cf. the contributions in Hickey 2010), but it is inadequate for the reconstruction of e.g. phonological systems. The situation is more difficult when it comes to spoken material since there are comparatively few early recordings of post-/colonial Englishes (again, the situation is different for Inner Circle varieties, where oral history projects and similar sources provide valuable documentation).

Focusing on the variety of English spoken in Ghana, the former British West African Gold Coast colony, the purpose of this paper is to explore some ways in which early popular music recordings might be used to reconstruct the phonology (and partly also morphology and syntax) of colonial and post-colonial Englishes in a situation where other recordings are near-absent.

Highlife was the epitome of West African dance music around the time of Ghana's independence in 1957. Music archives and collections around the world hold the gramophone and vinyl records from these early days, e.g., the African Music Archive (Mainz, Germany), the Bokoor African Popular Music Archives Foundation (Accra, Ghana), the Gramophone Records Museum of Ghana (Cape Coast, Ghana), and the British Library Archival Sound Recordings (London, UK). Recently, the Sound Archive of the Ghana Broadcasting Company has also been digitalized (cf. Coester 2010). The lyrics of some of these recordings represent authentic Ghanaian English from a time where hardly any other records of spoken language exist.

We will present our first results of the analysis of the Highlife lyrics collected, digitized and transcribed so far, including acoustic and auditive analyses of vowel subsystems where present-day Ghanaian English shows variation. For example, the STRUT lexical set is today variously realized with three different vowels in Ghanaian English, /a, e, ɔ/ (Huber 2004: 849) and it will be interesting to see in how far this variation can be traced back to the Ghanaian English of 50 years ago. We will also discuss methodological and technical problems connected with the linguistic analysis of early music recordings and discuss the implications of our results for the development of Ghanaian English in particular and for developmental models of varieties of English in general. We believe that diachronic studies based on recordings and focussing on language structure can make a significant contribution to models such as Schneider's (2003, 2007) 5-stage evolutionary model, that have for the most part been based on external language history and synchronic structural data.

Progressive aspect in Nigerian English

Robert Fuchs and Ulrike Gut (University of Augsburg)

Many authors have claimed that a non-standard usage of the progressive aspect is typical of new varieties of English such as Indian English, Ghanaian English and Singapore English (Platt, Weber & Ho 1984: 72f., Trudgill & Hannah 2008: 107, 130). Especially the usage of the progressive aspect with stative verbs in utterances such as “*Some people are thinking it’s a bad job*” (Sharma 2009: 181) has been named as a candidate for a universal structure shared by all of the new Englishes (Kortmann & Szmrecsanyi 2004: 1189, Mesthrie 2008: 626)

The few empirical studies that have so far investigated the usage of the progressive aspect in the New Englishes have found systematic meaning extensions such as the occurrence of the progressive with stative verbs, with verbs that denote ongoing changes as well as with verbs that refer to iterative unbounded situations (Sharma 2009, Balasubramaniam 2009, Van Rooy 2006). These have been explained with mother tongue influence or the specific substrate-superstrate interplay. Previous analyses, however, have been restricted to small data sets, to either basilectal speakers and conversational data (Sharma 2009) or student essays (Van Rooy 2006). Data from educated adult users has been neglected so far. Equally, stylistic variation of progressive aspect usage has not been investigated yet in the New Englishes although this is well documented for British and American English (e.g. Leech et al. 2009, Biber et al. 1999).

The aim of this study is to explore the system of progressive aspect marking in adult speakers of Nigerian English with a corpus-based methodology and to investigate variation in progressive aspect usage with register and style. The data was drawn from the ICE Nigeria, which is being compiled at the University of Augsburg (Wunder et al. 2010) and currently comprises 532,893 words of written and spoken language. A total of 1,994 progressive constructions were analysed. For comparison, the ICE-GB corpus was queried using IceCup 3.1.

The results show a distinct stylistic variation of the usage of progressives in Nigerian English with a frequency of progressives that is ten times higher in informal text types such as social letters and conversations than in formal ones such as academic writing. Equally, the incidence of rare progressive constructions varies with text type: Progressive passive constructions occur far more often in academic writing and news reports, text types that involve a formal and detached way of presenting information, than in private conversations and social letters.

The corpus-based study further demonstrates that the frequency of non-standard usage of the progressive in Nigerian English is quite low at less than 5% of all progressive constructions. It includes the usage of progressives with verbs referring to habitual nonbounded activities, with stative verbs referring to a mental state and with non-agentive stative verbs such as *hear*. Non-standard usage of the progressive is far more frequent in the informal text types social letters and private conversation than in academic writing. In comparison to British English (Leech et al. 2009), Nigerian English has a higher incidence of passive progressives, present perfect progressives and modal auxiliary+progressive constructions, especially in written language. Conversely, part progressives are rarer in Nigerian English than in British English. Sharma’s (2009) hypothesis on the specific substrate-superstrate interplay regarding progressive marking was only partly confirmed when tested against the Nigerians speakers with Igbo and Yoruba as first language.

Asian and Pacific English

Madam or aunty jee: **contrasting forms of address in British and Indian English(es)**

Tatiana Larina (Moscow, Russia)

Svetlana Kurtes (Portsmouth, UK)

Neelakshi Suryanarayan (New Delhi, India)

The paper reports on some findings of an ongoing research project investigating how cultural differences impact on cultural values, social categories and, ultimately, communicative styles in the British and Indian varieties of English. Here we shall focus on forms of address.

Since Brown & Gilman (1960), numerous studies emerged, investigating forms of address from a variety of frameworks and paradigms (e.g. Afful 2006; Aliakbari & Toni 2008; Carreira 2007; Clyne 2009; Ilie 2007, to name but a few more recent ones). Among other findings, they have shown that forms of address as a category are preeminently culture specific, the choice of which is conditioned by a number of variables such as social milieu, norms, beliefs, values, etc.

More specifically, the choice of an address form primarily depends on the actual communicative situation, but also on specific social and cultural contexts, indicating how interlocutors are connected or related to each other. When choosing a particular address form, participants in the communicative situation reflect such cultural dimensions as Social Distance and Power Distance (Hofstede 1984).

The paper will present some key differences between the two varieties of English – British and Indian – related to forms of address used in everyday interaction, taking into account characteristics of social and interpersonal relations, cultural values, politeness strategies and the categorization of reality which are particularly culture specific. Taking into account the rich cultural diversity of India, we have limited ourselves to those forms of address used exclusively in the northern states of India (as used by speakers of Indian English, as well as Hindi). Our initial premise was that the speakers of Indian English, in choosing address forms in their daily interaction demonstrate specific tendencies reflecting their native culture, since they belong to a culture with shorter Social Distance and higher Power Distance, unlike the speakers of British English, whose native culture is characterized by considerable horizontal distance and a rather low level of vertical distance. Some of the observed tendencies were:

- a) Frequent use of formal forms of address (*sir, ma'am, miss*), as well as the specific Hindi honorifics *ji (jii, jee) (madam ji, driver ji, waiter ji)* and *sahib (saab, sahib, sahib) (doctor sahib)*, when addressing strangers to show respect;
- b) Frequent use of family terms (*uncle, aunty, brother*, etc) when addressing strangers to show solidarity and closeness;
- c) Frequent use of Hindi family terms (*behen* 'sister', *didi* 'older sister', *bhaiya* 'brother', etc) when addressing strangers, demonstrating solidarity and closeness;
- d) Implementation of two opposite strategies: showing both solidarity and respect by using a family term combined with the traditional Hindi markers of respect *jee* or *sahib (aunty jee, uncle jee, behen jee, bhaii sahib or bhaisaab, doctor sahib)*.

The analyzed data was collected through ethnographic observation, questionnaires and interviews.

Doubling in New Englishes: Telling apart reduplication and lack of fluency

Manfred Sailer (University of Göttingen)

Corpora reflect language use and, as such, include both performance and competence aspects of language. It is, however, not always easy to separate these two. I will present a corpus study on word doubling in Singapore English and Hong Kong English based on the respective parts of the International Corpus of English (ICE). Doubling is seen as the adjacent repetition of the same word. Doubling can have several reasons, mirroring the competence-performance dichotomy: First, it can be an instance of total reduplication, i.e. semantically or pragmatically meaningful to express plurality, intensity, repetition, etc. Second, it can be an indication of lack of fluency, a way to fill a pause in speech. I will call these two types *reduplication* and *pause filling*.

Total reduplication is very restricted in British English (Quirk et al. 1985, Stefanowitsch 2007), but it is one of the innovations documented for many New Englishes (Wong 2004, Wee 2008 on Singapore English). Stolz 2008 argues that total reduplication is a universally available process which is present in first language acquisition and creolization, but absent in second language acquisition and pidgins. Given a differentiated picture of New English varieties as expressed in the Dynamic Model of Schneider 2003, 2007, I expect that reduplication is present in varieties that are in an advanced stage 3 (nativization) and higher in the model, but absent in varieties that are in earlier stages. On the other hand, the lack of fluency is a clear indication of performance difficulty and missing native-like competence. So, we expect pause filling to be more prominent in less advanced varieties (at least in the indigenous strand), whereas advanced varieties should show a rate of pause-filling doubling similar to that of native varieties. Schneider 2003 places Singapore English at stage 4 (endonormative stabilization) and Hong Kong English at an early stage 3 (nativization). This leads us to the following hypotheses: (H1) There is more reduplication in Singapore English than in either British English or Hong Kong English. (H2) There is more pause filling in Hong Kong English than in either British English or Singapore English.

I used the British, Singapore, and Hong Kong parts of the ICE (ICE-GB, ICE-SG, ICE-HK) to test these hypotheses. I applied the algorithm from Stefanowitsch 2007 to extract all instances of word doubling. This resulted in: ICE-GB (6,651 tokens; 744 types; 452 hapax legomena); ICE-SG (6,094 tokens; 924 types; 586 hapax legomena); ICE-HK (17,039 tokens; 1,305 types; 834 hapax legomena). Since the plain data do not distinguish between reduplication and pause filling, the challenge is to tear these two apart. The token numbers seem to contradict H1: There is no statistical difference between GB and SG, but HK shows a significantly larger amount of doubling than GB and SG. I calculated strict and global productivity (Baayen/Lieber 1991, Plag 1999): Doubling is most productive in ICE-SG and equally productive in ICE-GB and ICE-HK. Thus, productivity helps us to identify reduplication as a grammatical process in Singapore English, but less so in British and Hong Kong English, confirming H1.

To explain the high number of doubling tokens in ICE-HK I looked at what words are doubled. In ICE-HK these include many instances of non-lexical pause fillers such as *uhm*: *uhm*-doubling covers 18.61% of the tokens in ICE-HK, 8.99% in ICE-GB and 6.51% in ICESG.

This supports H2, that pause-filling doubling is most prominent in Hong Kong English, but there is no difference in fluency between Singapore English and British English. The extracted data on doubling do not discriminate between competence (reduplication) and performance (pause-filling). However, applying the methods of quantitative morphology we reach an empirical confirmation of the predictions on fluency and grammatical innovation in New English varieties based on the Dynamic Model.

Orthographic Developments

Tracing orthographic change in corpora: A methodological approach to the study of English compound spelling Christina Sanchez-Stockhammer (University of Erlangen-Nuremberg)

Writing is generally regarded as a system which is secondary to spoken language (de Saussure 1916/2005: 45). This may be the reason why spelling appears to constitute a particularly arbitrary aspect of language. Nonetheless, a more or less conventional standard orthography seems to have evolved in the course of the history of the English language in spite of the fact that no single institution is endowed with the authority to impose orthographic rules on the users of English. A closer look at the apparently uniform and seemingly unquestioningly accepted spelling system will reveal, however, that there are still considerable discrepancies in certain areas. My paper will introduce the methodology of a comprehensive study on one such area, namely the spelling of English compounds. It is the aim of this study to discover whether the preference for particular spelling variants is linked to certain word-immanent features and what reasons may underlie divided usage.

In the past, orthographic issues were approached from a purely prescriptive point of view, with the rules being formulated first and usage expected to follow. The advent of descriptive linguistics in the 20th century changed the methods, and linguistically oriented current style manuals attempt to provide a description of the language rather than a summary of prescriptive rules (e.g. Peters 2004). The new methods of investigation take into account usage and are based upon the empirical study of actual language data, e.g. in the form of corpora. However, corpus research always presupposes that the researcher determine some kind of expected pattern in order to obtain any results. This makes the study of orthographic changes particularly difficult because these are often idiosyncratic, so that we do not know in advance what precisely to look for. My paper will discuss the opportunities and limitations of corpus-based orthographic studies, e.g. the problem that systematic tolerance of variation in one or more places in an orthographic word often results in unwanted hits (e.g. *latter* and *better* for *letter*).

The spelling of English compounds, however, is an orthographic issue lending itself to corpus-based empirical research because the expected patterns are usually of three kinds only: open (e.g. *key ring*), hyphenated (e.g. *fat-free*) or closed (e.g. *weekend*). This particularity allows me to test a common hypothesis, namely that compound spelling usually develops from the open to the closed mode (cf. e.g. Quirk et al. 1985: 1537), in a fairly automated way. The paper will give an outline of how this is done. The consideration of the 20th century for the diachronic study has the advantage of permitting the prediction of future developments based on selected corpora of the Brown family – e.g. BLOB-1931, LOB (1961), FLOB (1991) and BE2006 for British English.

My paper will show how we can use the above corpora to discover changes in the orthographic system and which theoretical issues need to be settled first. An important point for discussion will be how to determine which items to use as the basis for the study of English compound spelling, i.e. how to decide what can be considered a compound. Several criteria have been suggested in the literature (e.g. Bauer 1998) to distinguish compounds from phrases; among them fore-stress, positional mobility, reference to a single idea and orthographic unity. That the latter cannot be considered a watertight criterion – at least not for English compounds – should be obvious from my research question already. However, counterarguments can be found for all other criteria as well, thus reducing their reliability. In addition, compounds need to be set apart from other types of lexeme, particularly from derivatives. I will therefore argue in favour of an approach that uses material from dictionaries as the starting point and then goes through several stages of selection (e.g. coding as compounds by lexicographers, occurrence in a predetermined number of reference works, presence of at least two free bases and no affixes on the highest level of analysis). As the consultation of reference works poses the danger of circular reasoning because established compounds might share certain qualities that make them different from non-listed compounds, the dictionary-based research will be complemented by a small-scale study which takes into account any compounds occurring in a number of real-language texts. This will enable me to test whether the results obtained for the core of the compound category also apply to its periphery.

Canadian English

Ethnodialectology: Ethnicity and regional variation in Canadian English

Charles Boberg (McGill University)

It has not so far been possible to define precisely the eastern border of the principal Canadian English dialect region because of the uncertain status of Montreal, once Canada's largest city and, despite the minority status of its English-speaking population, still an important urban concentration of English-speakers. Boberg (2005b) found that Montreal English is distinct from other Canadian varieties at the lexical level, a status first established by Hamilton (1958). More recently, analysis of regional variation in Canadian English has focused on phonetic variables. At this level, the status of Montreal is not clear. Whereas analyses of the Canadian Vowel Shift (Clarke, Elms and Youssef 1995; Boberg 2005a, 2008; Labov, Ash and Boberg 2006: 224) have included Montreal within its territory, research on Canadian Raising has suggested that Montreal is an exception to the pattern found in other Canadian cities (Hung, Davison and Chambers 1993), while the *Atlas of North American English* excludes Montreal from several other isoglosses that define Inland Canada (Labov, Ash and Boberg 2006: Map 15.7).

This paper examines the taxonomic status of Montreal in much greater detail than was hitherto possible, by means of a fresh methodological approach. It presents acoustic phonetic data on vowel production by 93 native speakers of Montreal English from three ethnic groups, both sexes and a broad range of ages and social backgrounds, compared with identical data from 73 speakers of other varieties of Canadian English. This approach allows regional comparisons to be made while controlling for speakers' ethnic origins, an important analytical step given the ethno-phonetic variation in Montreal English identified by Boberg (2004; 2010: 222).

While Montreal and Canadian English are identical in one important respect -- the low-back merger -- they are distinct in several others, including structurally important issues of allophonic variation and even phonemic contrast. In particular, Montreal shows only variable presence of Canadian Raising and a complete absence of the merger of /æ/ and /e/ before /r/ (in *marry* and *merry*), features found everywhere else in Canada. Finally, the allophonic organization of /æ/ is simpler in Montreal, with very little raising before nasals and /g/. However, many of these differences are considerably reduced when only the British-ancestry population of Montreal is examined: the greatest divergence from Canadian English is found among Italian-origin Montrealers. The data demonstrate that Montreal is one city where ethnic origin, even within the European-ancestry population, cannot be ignored in making regional comparisons.

**New-Dialect Formation cum Dynamic Model:
Language attitudes and the case of Vancouver English**
Stefan Dollinger (University of British Columbia at Vancouver)

New-Dialect Formation (NDF) theory has argued that the initial dialect mixing process in colonial settings can be described and, within limits, even predicted, based on the numerical strength of the input varieties (Trudgill et al. 2000, Trudgill 2004). The Dynamic Model (DM) is often considered the antithesis of the NDF, as it focuses on the “identity-driven process of linguistic convergence” (Schneider 2007: 30, Schneider 2003) that NDF denies for some stages of the process. While NDF remains contested (see Discussion [of New-Dialect Formation] 2008, Schreier 2008: 54), the homogeneity of linguistic systems in the New World, including Canadian English, seem to confirm, overall, its methodological base. This paper aims to test the applicabilities of NDF and DM in the Vancouver setting using historical and present-day data, which include the changing attitudes in the Vancouver population over the last 30 years.

Vancouver English is a fairly young variety. Founded in 1886, the City of Vancouver and its conurbation “Metro Vancouver” has undergone the rapid growth that is characteristic of major west coast cities. The short time-span since its original settlement allows the study of the variety from two vantage points. First, reconstructions of input varieties and earlier stages of the variety, as historical sources abound, are feasible and, second, synchronic studies in apparent time are able to inform models of new-dialect formation, as they reach back to the third generation of settlers. The present paper aims to show how perceptual dialectology in apparent time, through self-reports of speakers' attitudes to Vancouver English (Chambers 1994), can inform discussions of dialect mixing and help establish the overlap between NDF and the DM.

Responses to a questionnaire on Language Attitudes from 427 Vancouverites in the fall of 2009 show a marked contrast with earlier findings of low awareness of a Canadian linguistic identity (Warkentyne 1983, Gulden [Halford] 1979). Recent studies on language and ethnicity in other Canadian contexts show only modest dialect diversification, and not yet at the levels expected by the DM (Schneider 2007: 247-50). The language awareness data suggest that Vancouver English is at stage IV, “Endonormative Stabilization”, in the DM, which is equivalent with “focusing” in the New-Dialect Formation model, and not at Stage V, “Diversification”. In CanE, the second and third immigrant generations largely share the same linguistic system with Standard Canadian English (e.g. Hoffman 2010: 135, Hoffman and Walker 2010: 59), Montreal being an exception (Boberg 2004: 563). The data suggest that NDF and DM are not mutually exclusive, but may perhaps be fruitfully combined into one coherent theoretical approach. While such endeavour proper is beyond the present paper's scope (see Hickey 2003 for some thoughts), data indicating common ground between NDF and DM suggest that it is not “NDF vs. DM”, but “NDF cum DM”.

Irish English

Detecting historical continuity in modern Singapore English:

A case study of the present perfect

Julia Davydova (University of Hamburg)

It is a truism that New Englishes are shaped by phenomena such as substrate influence (transfer) from the native language, various forms of superstratal input (Standard English as a target language in the classroom setting, historical input varieties of various kinds) and universals of second language acquisition. The major goal of this study is to explore the issue of the contact-induced influence exerted by the historical variety of Irish English (IrEng) on the formation of the present-day Singapore English (SingEng). A mention of the existence of such a contact is documented in Schneider (2007) and earlier in Gupta (1998).

Relying on the model and method employed in variationist sociolinguistics, the study provides empirical substantiation for the claim that there is indeed historical continuity between New Englishes and earlier forms of English (cf. Mesthrie and Bhatt 2008: 188ff.). More specifically, the study demonstrates that while transforming patterns of variation in the process of language contact, second-language learners appear to be sensitive to cognitively salient patterns of variation invariably reproducing them in their speech.

This talk is organised as follows. Having briefly outlined the major terminological distinctions drawn in the study, we proceed to presenting a general framework for studying varieties in contact. Having presented the data and methods employed in the study as well as its results, we proceed to the discussion of findings in the light of the variationist model proposed in Meyerhoff (2009). We conclude that (i) while effects of language contact manifest themselves in a shift of variation patterns, some patterns of variation remain stable across varieties in contact arguably as a result of their cognitive prominence and (ii) variationist methodology is an excellent tool for ascertaining the degree of impact of one variety on another.

The study draws on data obtained from the *Hamburg Corpus of Irish Emigrant Letters* (HCIEL), a historical corpus representing a variety of the 19th-century Irish English. The SingEng data was obtained from the Singaporean component of the International Corpus of English (ICE). In order to enlarge the scope of our comparison, a London-Lund Corpus (LLC) of spoken English was included into the analysis. The corpus represents a mainstream variety of English (henceforth, StEngEng) spoken by native educated speakers in England and provides a useful yardstick against which the non-native and historical data can be measured and compared. Serving as a reference variety, StEngEng enables the analyst to ascertain which patterns of variability of the present perfect attested in SingEng can account for its close affinity with the standard grammar of English and which patterns are indeed to be attributed to the historical input from IrEng.

**“I may tell you”:
A preparation of news to come in 19th century Irish English emigrant letters**
Marije van Hattum (University of Manchester)

“I may tell you something that perhaps your [sic] would like to here[sic]” (OC-Dev 2). In Irish English emigrant letters the modal verb MAY followed by the main verb TELL is sometimes used to introduce news to the family. In the 18th and 19th centuries it was not uncommon in other varieties of English to introduce a proposition by using a modal verb + *tell*, but Irish English is different from other varieties of English spoken in Britain. Irish English emigrant letters make significantly more use of this construction and it is the only variety of English that can have MAY in the slot of the modal verb, as opposed to other verbs such as MUST, CAN, HAVE TO, etc.

The first aim of this paper is to attempt to investigate the status and uses of the construction with a modal verb + *tell* from a pragmatic point of view. The second aim is to investigate the origins of the construction with MAY and to try to explain why *may + tell* is only used in Irish English. This paper only deals with those instances of *modal + tell* where the subject is in the first person and the indirect object in the second person. Furthermore, the news that is being conveyed must be within the same paragraph, so that *modal + tell* can be seen as an introduction to or preparation for the news and not a promise of telling it in the future. These features are exemplified in (1).

- (1) I must also tell you that John graham bought[sic] a farm. (ML-Bro2)
 1st nom modal tell 2nd acc news

The data come from Fitzpatrick's (1994) corpus of 19th-century emigrant letters and from the Corpus of Irish English Correspondence (McCafferty & Amador-Moreno, under construction). In total I have examined approximately 160,000 words in 285 letters written by male and female emigrants from various social classes (middle and lower) and geographical areas (north, east and west).

The Invited Inferencing Theory of Semantic Change (IITSC) has been successfully applied to the study of modality and discourse markers in both English and Japanese (Traugott & Dasher 2002), which therefore makes it a suitable framework for the analysis of the modal + tell construction. Traugott and Dasher (2002: 35) argue that “there is a path from coded meanings to utterance-token meanings (IINs) [Invited Inference meanings] to utterance-type, pragmatically polysemous meanings (GIINs) [General Invited Inference meanings] to new semantically polysemous (coded) meanings”.

This study will explore the possibility of the *modal + tell* construction having undergone these stages of IITSC and become a discourse marker used to introduce news in personal letters in 19th-century Irish English. Though the meanings of the modal verbs in these constructions have become obscure, they are not lost. The modal verbs are not interchangeable and each modal verb has its own function as a possible discourse marker, retaining hints of its ‘core’ meaning. In the specific case of *may + tell*, I will argue that this ‘core’ meaning was not ‘permission’, as would be expected, but the older meaning of ‘have power to’. Thus it is hoped that the results will shed new light on the development of modal verbs in Irish English in general, and of MAY in particular.

Invariant Tags in Irish English: A Corpus based Analysis

Stephen Lucek (Trinity College Dublin)

Since the pioneering work of Schiffrin (1987), the study of discourse markers has progressed considerably. Discourse markers contribute to understanding discourse coherence, defined by Schiffrin as “how speakers and hearers jointly integrate forms, meanings and actions to make overall sense of what is said” (Schiffrin (1987, p. 49)). Discourse markers are a vital element to the coagulation of information packets, adding colour and variety to the discourse coherence. A great deal of work has been done on discourse markers in Irish English, most notably Amador Moreno (2005) and Kallen (2006).

Kallen (2006) goes deeper into literary tradition to find examples of discourse markers in the presentation of Irishness from the 17th Century onward. He suggests that the earliest examples “are etymologically derived from Irish and carry over similar discourse functions” (Kallen (2006, p. 1)).

While the above-mentioned studies have covered discourse-markers in general, the current study will deal with specific utterance-final discourse markers: Invariant Tags. Invariant Tags themselves are far less robust than utterance-medial discourse markers and have not been described in nearly as much detail as discourse markers in general. Much has been written about Invariant Tags. Columbus (2009) investigated the frequencies and selection of Invariant Tags in five varieties of English. This study is of paramount importance to the current study as it also relies on International Corpus of English (ICE) corpora (see ICE Project, below). Berland (1997) looked at Invariant Tag use by teenagers in London. Andersen (1997, 1998) used the same database (Corpus of London Teenage Speech (COLT)) as Berland (1997) to investigate the nature of Invariant Tags and whether certain types of tags can be considered Invariant.

Some Invariant Tags are indicative of the varieties of English in which they are most prevalent (e.g. Canadian ‘Eh’, Singapore English ‘Lah’). As Invariant Tags are a natural occurrence of conversation and are only present in literature when describing speech, data collection methods must be adjusted to describe them.

This study investigates what exactly Invariant Tags are, which Invariant Tags occur in Irish English, in which situations do they occur, how Irish English use of Invariant Tags compares to that of other varieties of English worldwide, can Invariant Tag use be a comparison tool for World Englishes, and what can the extant models of World English tell us about different types of English? Using the SPICE-Ireland corpus of Irish English, a qualitative analysis of Invariant Tags extends the existing definition of Invariant Tags and describes the functions of Invariant Tags. Finally, a quantitative analysis of the Invariant Tags of Irish English are compared with those of the five varieties of English described in Columbus (2009) with conclusions drawn from that comparison.

Phonological Topics in American English and New Englishes

Flaps and other variants of /t/ in American English: Allophonic distribution without constraints, rules, or abstractions

David Eddington (Brigham Young University)

The distribution of the flap allophone [ɾ] of American English, along with the other allophones of /t/, [t^h, t̄, ʔ, t] has been accounted for in various formal frameworks by assuming entities and mechanisms such as prosodic licensing, syllable boundaries, foot structure, nonsurface apparent phonetic features, rules, constraints, and derivations. The desirability or usefulness of these entities and mechanisms within a formal framework is not at issue in the present paper. Instead, a computationally explicit model of linguistic performance is used (Skousen 1989, 1992) in order to account for the distribution of the allophones of /t/ without recourse to such entities. On the one hand, the psychological reality of such abstract entities in actual language processing is highly questionable. On the other hand, the simulations that were carried out suggest such abstractions are not needed because analogy to surface apparent variables such as phones and word boundaries is sufficient.

In analogy, the particular allophone of /t/ (i.e. [ɾ, t^h, t̄, ʔ, t]) that appears in a given context is determined on the basis of similarity to stored exemplars in the mental lexicon. From an acquisitional standpoint, analogy to stored exemplars dispenses with the need for rule induction. Analogy also explains the stochastic nature of linguistic performance. In the present study, 3,738 tokens of the allophones of the phoneme /t/ were extracted from the TIMIT corpus and constitute the database from which analogs were chosen. The variables used included the three phones or boundaries on either side of /t/, the stress of the syllables preceding and following /t/, and whether a syllable boundary appears next to /t/. The model proves quite successful in predicting the correct allophone, and the errors made are generally possible alternative pronunciations (e.g. moun[ʔ]ain, moun[t^h]ain). The success rate changes little when only small sub-samples of the database are incorporated. In addition, exemplar-modeling is found to be quite robust because even when features such as syllable boundaries and stress are eliminated (both of which are critical in rule approaches), allophony is still highly predictable.

New Englishes and the Emergence of the Unmarked

Caroline Wiltshire (University of Florida)

Phonologists have a longstanding interest in how pidgins, creoles, and vernaculars share properties, regardless of their substrate, leading to discussion of which properties can be claimed to unmarked/default settings of the language faculty (e.g., Filppula, Klemola and Paulasto 2009). Similarly, SLA research suggests the unmarked is revealed in the patterns of acquisition of marked structures by learners whose L1 lacks such structures (Broselow et al. 1998). With the recent increase of detailed phonetic and phonological descriptions of new and emerging Englishes (e.g. Kortmann and Schneider 2004), we are now in a position to begin evaluating hypotheses about the emergence of the unmarked in new Englishes, both outer and expanding circle. Such descriptions includes work on acoustic phonetics of vowel systems (Deterding 2003, Sarmah et al 2009, etc.), both of the English and of the substrate/L1 languages, and detailed descriptions of consonant reductions (Bobda 2007, Wiltshire 2005).

Phonological markedness has been applied both to segments in an inventory (paradigmatic) and to sequences of segments in larger structures, such as syllables (syntagmatic), and has been based on evidence from cross-linguistic typology, phonetic simplicity, and implicational relationships (Trubetzkoy 1969, Jakobson 1968, Maddieson 1984, and ff). In this paper, I examine both vowel contrasts and the reduction of consonant clusters in a variety of new Englishes. In each case, New English varieties take steps towards reducing markedness found in the target English, although in some cases potential transfer from the substrate complicates the interpretation of these reductions as purely markedness driven.

The vowel system of inner circle Englishes is recognized as relatively marked in several ways, for example containing a contrast in tense/lax high vowels [i:]/[ɪ] and [u:]/[ʊ] and multiple mid-vowels. Studies of the acoustic phonetics of the vowel systems of New Englishes reveal that they often contain fewer contrasts than those of native Englishes, and that the choice of which vowels/contrasts are available reflects both the substrates and markedness constraints (e.g., Arabic English: Munro 1993, Brunei English: Sharbawi 2006, Singapore English: Deterding 2003, Thai English: Sarmah et al. 2009).

The treatment of consonants and clusters gives clearer evidence for markedness at work. Syntagmatic markedness in the form of complex consonant clusters in codas is also a noticeable feature of native Englishes. When the substrate lacks the complexity of the target English, the New English follows markedness scales in determining which consonants/clusters appear (e.g., Chinese English: Broselow et al 1998, Tibeto-Burman India English: Wiltshire 2005, African Englishes: Bobda 2007). For example, clusters more frequently produced correctly tend to be less marked in terms of the sonority sequencing generalization (Selkirk 1984), which requires consonant clusters in the coda to fall in sonority towards the end of the syllable, and the minimal sonority distance requirements (Steriade 1982). Clusters with consonants that are too close in sonority or that violate sequencing are more likely to be reduced. As native varieties also have cluster reductions, it could be that differences are in degree and not in kind (Schreier 2009); we find more reductions when comparing inner to expanding circle, just as we find more when comparing formal to casual style, or before consonants than before vowels. The data reveal deletions based on cross-linguistic markedness, within each variety, style, or phonetic context.

Thus, this paper explores to what extent the descriptive data compiled from new Englishes tends to support the claim that these new English varieties reveal the emergence of the unmarked, finding that when markedness can be separated from substrate influence, new Englishes indeed show a preference for the unmarked.

Phonetically accidental and systematic gaps: Exemplifying British English triphthong reduction and American English alveolar sequences

Toshihiro Oda (Fukuoka University)

Assuming that nonexisting segments consist of possible ones (accidental gap) and impossible ones (systematic gap) in the phonological system of a language, the two pieces of previous research have posited two levels of the gaps: lexically (Chomsky and Halle 1965) and phonologically (Iverson and Salmons 2005) accidental and systematic gaps. The former specifies, on the basis of phonotactic constraints, that, given the word *brick*, the lack of *blick* is assumed to be accidental, while that of *bnick* systematic. In the latter, tense vowel plus ESH was lacking in early English, but the gap has been filled in the later history. The sequence is allowed to occur in the grammar and easy to overcome for native speakers, and thus constitutes accidental gap. Building on phonetically driven phonological theory (Licensing by Cue by Steriade 1997, Inductive Grounding by Hayes 1999, Bidirectional Phonology and Phonetics by Boersma 2007), the paper presents the segmental gaps caused by physical activity.

Phonetically accidental and systematic gaps are based on a certain physical activity, that is to say, lack of a movement, lack of an air flow, etc. Phonetic bases hold to segments with different syllabic affiliations (onset, nucleus, coda) and with different articulators (e.g. dental, alveolar, palato-alveolar, or front, mid, back), in the sense of identical phonetic category. On the condition that an articulation is easy or difficult for speakers, the articulation is required to hold to all of the segments with it irrespective of the syllabic slots and the articulators. Phonetically preferred patterns might not exist in earlier stage, but they appear with allophonic shift and the gap is filled. The articulation conforms to the one of the language and had been highly possible to exist in it prior to the appearance. Regarding the syllabic slots and the articulators, to fill the gap proceeds by way of more than one; identical articulation appears, say, at dental, palate-alveolar and uvular. This is referred to as phonetically accidental gap. Phonetically systematic gap is equivalent to the corresponding ones in the previous research in the sense of difficulty to overcome and no account for the existence. The feature on the former is that phonetically dispreferred patterns do not appear in most or all of the languages. The reason for it stems from the universality that the lack of certain difficult articulations is shared by most or all. This gap is peculiar to consonant sequences due to some extremely impossible ones relative to sequences of vowels. The sequences of *tap/flap + alveolar and *dental/palato-alveolar + alveolar are involved in the crosslinguistically unpronounceable examples which are accounted for by physical activity.

BrE triphthongs undergo diphthongization due to the loss of the second elements (e.g. power, fire). For the reduced [aə], the first low mid vowel shifts to the second central mid and, given cardinal vowels, the distance between the two elements is 1.5. The reduced diphthong commonly occurs word-finally. Non-word-finally, it is followed by the alveolars (science, Howard), which do not have the slight effect to shift to front. Both the mid distance between the two elements and the slightly fronted second element are involved in the British preferences. The diphthong at issue did not exist in earlier stage, but the gap has been filled in recent Received Pronunciation (cf. Wells 1997). Thus, it means phonetically accidental.

Despite the highly common tap, AmE has the syllabic [n] preceded by either nasally released /t/ or glottal stop (eaten, button). The lack of the articulation such as the tap plus the syllabic [n] represents phonetically systematic gap. Differently from the laterally released tap, the tip of the tongue returns to the neutral position, if pronounced, before syllabic [n]. The appearance of the cluster [nr] phonetically differs from *[r̥n], the sonority contours being similar, since the former has, without the return, the firm contact of the tip with alveolar. Relevantly, alveolar stops neither precede nor follow the tap, whereas other rhotics precede an alveolar stop in the rhyme. This also stems from the return to the neutral position. There is no AmE articulation involving two alveolars of this sort without a vowel intervened or with a vowel deleted.

Academic Styles

Towards the emergence of technolectal Nigerian English

Presley Ifukor (University of Osnabrück)

Since 2000 there have been progressive developments in the Nigerian telecommunications and information technology sectors, and the truth is that Nigeria is witnessing a mobile telephony and Internet boom. According to the Nigerian Communications Commission (NCC) in its December 2009 report, there are 73 million active mobile phone subscribers in Nigeria which indicates that Nigeria has become Africa's largest mobile phone market. Statistically, the Nigerian mobile market grew exponentially from 422, 000 in 2001 to 73 million in 2009 which translates to a progression from 0.33% of the population in 2001 to 48.7% in 2009 (Pyramid Research, 2010). As a consequence of the increased number of mobile phone subscribers, short message service (SMS) text messaging has become an integral part of interpersonal communication in Nigeria with noticeable sociocultural flavour (Awonusi, 2004; 2010; Chiluba, 2008; Elvis, 2009; Taiwo, 2008; 2010a). Text messaging also features prominently in participatory politics just as mobile telephony is crucial to democratic activism (Obadare, 2006). In fact, Nigerian textizens used the SMS feature of mobile telephony to monitor the 2007 Nigerian elections. As documented in the report of the Network of Mobile Election Monitors (NMEM, 2007), an estimated 11,000 text messages were received from ordinary Nigerians on April 21, 2007 the day of the presidential election (Banks, 2007). NMEM used a free FrontlineSMS platform to collate all the text messages. Through this mobile service, thousands of ordinarily faceless and voiceless Nigerian masses found a medium to make their voices heard and to express their candid opinions about how the presidential election was conducted. The mobile technology therefore enabled the electorates to also serve as volunteer local observers. Apart from politics and civil liberty activism, mobile phone plays an instrumental role in the private and public affairs of ordinary Nigerians than generally acknowledged. Similar to the growth in mobile telephony services and subscription, Internet connectivity in Nigeria is increasing and improving. From a mere 200 thousand Internet users in Year 2000, there are now 43 million Nigerian Internet users (as of December 2009, according to ITU, 2010) which represents 28.4% of the nation's population. On the African continent, Nigerians account for 39.6% of Internet users, which is the highest. Most Nigerian Internet users access the Internet at work or Internet cafés (Sesan, 2010) and a growing population of private users has access at home via wireless connection and mobile phones at relatively affordable prices. Therefore, the penetration of mobile telephony and the Internet in Nigeria has techno-linguistic implications, some of which are as follows:

1. Being the most linguistically diversified African country in conjunction with the fact that Nigerians are embracing social media and mobile texting at a phenomenal rate (Ifukor, 2010), more linguistic data are being generated, which opens up a window of research opportunities for language scholars as language is at the centre of communication via mobile phones and the Internet. Moreover, Internet and mobile phone textual data are a goldmine of resources for diachronic and synchronic linguistic studies.

2. Following from the above, the diversity of languages and cultures in Nigeria will surely be reflected in the way Nigerians compose and transmit their messages via computers and mobile technologies (cf. Chilwa, 2010; Taiwo, 2010b). Therefore, researching the computer mediated communication (CMC) of Nigerians is worthwhile because of the strategic position Nigeria occupies in Africa as the most populous country whose domestication and nativization of the English language coupled with the multilingual setting in the country provide useful data for empirical investigation.
3. With more users of English and CMC technologies on the African continent situated in Nigeria, a subvariety of English influenced by the ubiquity of social and mobile technologies and the informality of the media will give rise to the emergence of what we call technolectal Nigerian English (TNE). This is hinged on the persuasion that globally available technologies can provide the impulse for a locally enacted symbolic creativity (Androutsopoulos and Scholz, 2002) in the form of localization of ICT and the globalisation of indigenous cultural practices and expressions. Thus, the underlying sociocultural manifestations of Nigerianisms in digital discourse are instantiations of transnational transcultural flow and fusion of ideologies (Pennycook, 2007) in global Englishes, whereby language responds to global trends.

Fusing aspects of Stewart's (1965), DeCamp's (1971) and Bickerton's (1973) continuum of lects, we propose a modification of the existing sociolinguistic Nigerian English continuum (cf. Awonusi, 1985; Jibril, 1982 and Okoro, 1992) along two dimensions, namely levels of literacy and technological savviness, to accommodate the embedding of the technolect. The modification also provides a paradigm for the description and characterization of this new sub-variety of English necessitated by globalisation. In terms of characterisation, TNE is a sublectal variety of English. It is a 21st Century post colonial phenomenon rooted in the globalisation of technological innovations in digital discourse and the localization of linguistic systems for the projection of indigenous ideologies, cultural practices and expressions. Using computational and online tools for data collection, the corpus for this study comprises mobile phone texts, emails, blogs, tweets, Facebook and online forum messages composed by Nigerians.

**The Phraseological Profile Model applied:
New insights into academic speech and writing**
Ute Römer (University of Michigan)

Starting from the observation that meaning does not primarily reside in individual words but in the phrase (see Sinclair 2004: 148), this paper focuses on the examination of recurring phrases in language. It introduces a new analytical model, the Phraseological Profile Model (short PP Model), that leads corpus researchers to a profile of the central phraseological items, their internal variation, their functions, and textual distribution (highlighting instances of textual colligation, Hoey 2005), in a selected text or text collection. The novelty of the approach lies in bringing existing corpus-analytic methods together in a new way while introducing a few new techniques and concepts along the way. The paper shows how Sinclair's (1996) search for units of meaning can be continued with more powerful software tools designed for phraseological analyses.

Since it is well known that there is a great deal of variation across registers (see, e.g., Biber 1988, Biber et al. 1999) and that meanings are expressed in different ways in different text types, it is suggested that the focus of our phraseological explorations be on subsets of language, or "restricted languages" in Firth's 1956/1968 sense. In this paper, the PP Model is applied to two corpora of different types and different sizes, both consisting of academic English and capturing part of the discourse of the global community of linguists: (1) a medium-sized (3.5-million word) corpus of online academic book reviews from the field of linguistics and (2) a small (100,000 word) corpus of transcripts from a specialized conference on applied linguistics and academic discourse analysis. Focusing on corpus methodology, this paper will demonstrate how the PP Model can help highlight central aspects of academic English and answer the questions "What are the most common phrases in the restricted language under analysis?", "How are meanings created in the discourse?" and "How are phraseological items related with text structure?" The results presented in this study may have important implications both for linguistic description and pedagogical practice.

Varieties of English in the classroom

Peter Siemund, Julia Davydova, and Georg Maier (University of Hamburg)

Varieties of English as well as the global expansion of the English language have been the focus of academic interest for many years. Consequently, a large body of scientific literature, corpora, and empirical studies has formed, covering a multitude of aspects, and yielding many new linguistic insights and generalisations.

To our knowledge, however, there has not yet been any systematic effort to refine this topic in such a way as to make it suitable for the needs of secondary schools (in EFL settings) and first-year university students with no or little linguistic background. In Germany, for example, the curricula for the school-leaving exams have been addressing English in different regions of the world for several years and will continue to do so in the future. Nonetheless, there are no linguistic materials suitable for students and teachers that address these issues in detail.

In our conference contribution, we would like to present the results of a threeyear research project carried out at the University of Hamburg in cooperation with more than a dozen EFL teachers, the Hamburg teacher training institute (Landesinstitut für Lehrerbildung und Schulentwicklung) and more than 300 German secondary school students, which resulted in the compilation and production of teaching materials with a strong didactic outlook that present an overview of different varieties of English and their characteristics.

Based on Braj Kachru's model of the three circles of English, the teaching materials provide a survey of different regional varieties of English, encompassing a discussion of Celtic Englishes, British English, American English, Australian English, Nigerian English, Indian English, South African English as well as some EFL varieties. The teaching materials pursue an empirical-inductive approach and are especially designed for advanced students of secondary education (in EFL classes) and for early undergraduate students of English. Students are familiarised with the characteristics of the different varieties of English using authentic audio, video, and text materials from these regions and are encouraged to discover the specifics of these varieties by themselves. Furthermore, a wide variety of exercises are included in the teaching materials ranging from comprehension, analysis, and text composition exercises to role-playing and language games designed to cater for the individual needs of the diverse teaching situations. Finally, the fact that the teaching materials were constantly tested, supervised, and improved during their genesis makes them unique.

In our contribution, we present a methodology to convert state-of-the-art research on varieties of English into a user-friendly classroom product. Thus, we do not only hope to open a highly fascinating research field to the student population but also to encourage future transfer of linguistic research results and knowledge from a small expert community into the wider public.

Letters and Literature

Same Tune, Different Songs:

Relevance Theory, Accountabilities, and Collocations in *Lord of the Flies* Criticisms

Dustin Grue (University of British Columbia)

This paper proposes to use principles of Relevance Theory (Sperber and Wilson 1986/1995) to identify features of *textual accountabilities* in the body of literary criticism of William Golding's (1954) *Lord of the Flies*. Since Sperber and Wilson proposed Relevance Theory (RT)—a pragmatic theory of context-sensitive communication—nearly 25 years ago, it has had massive uptake in a variety of academic disciplines: from translation theory (Mateo 2009; Dooley 2008; Gutt 1990) to poetics (Pilkington 2000; Uchida 1998), and—of course—pragmatics (Carston 2009; Blakemore 2008; Blakemore and Carston 2005; Sperber and Wilson 2005, 2002). So far, however, little progress has been made in developing a viable *method* for using relevance-theoretic principles in corpus investigations. I will propose a method for using this theory, a way in which it might be 'turned toward' a body of texts to illuminate the 'textual relationships' therein: how authors and text 'account', and remain 'accountable', in a corpus.

This corpus is comprised of the 74 English *Lord of the Flies* criticisms (2 books, 47 book chapters, and 25 journal articles) published between 1960 and 2009 listed in the Modern Language Association International Bibliography. I have compiled this corpus because it is substantial, but certainly not un-wieldy, and demonstrates a curious textual feature: numerous collocations are present within the corpus, formed from critics reiterating the same or similar phrases, to the point that this repetitive language becomes almost *idiomatic* in the discourse. And yet, while there is a certain degree of 'coalescence' of expression, there is a substantial divergence in the articles' critical purpose – to put it simply, critics use the same words to mean very different things. Since RT is particularly sensitive to identifying speaker's meaning by context and intent—and not just lexical realization—RT should be particularly adept at illuminating what is happening at this point of divergence, and revealing the *accountabilities* supporting these phrases.

I am here inspired by Charles Bazerman's (1988) dual-use of the term 'accountability'—to refer to both the ways in which 'text' can account for, and remain accountable to, its correlate 'experience'—in his analyses of scientific discourses. However, beyond accountabilities I hypothesize that collocations and idiomatic phrases can hold multiple relationships; in the present corpus, not just between text and the experience of criticism but between author, critical practice, and the discourse in which they participate.

Indeed, although it might be assumed that such highly stable, recurring phrases are indicative of authors establishing 'commonplaces' or utilizing a discourse-specific lexicon in the literature, preliminary findings suggest that these collocations recur as *paradigms*, in Giorgio Agamben's ("What is a Paradigm?", 2009) conception, in that they establish 'singular', instead of 'global', horizons of meaning. That is, rather than functioning as a method of 'grounding', or working as enthymemic references in a part whole relationship within the corpus, these phrases are utilized analogically in order to bring attention to points of divergence – by identifying and separating the text from the whole body of criticism that the author at once acknowledges and ignores.

As I have noted, while a comprehensive method for using RT in corpus research has yet to be fully developed, my approach to this method does follow from Janet Giltrow's (1994) work in which RT, breaks in textual cohesion (Halliday and Hasan 1976), and rhetorical genre theory were used to explain the unstated assumptions (un)announced within a corpus of legal documents. As well, I am using bibliometric principles in coordination with RT, as described by Howard White (2007a, 2007b), in order to limit the reliance on 'close reading' critical practices – an RT methodology remains viable especially as it might be used 'distantly' with a large collection of documents.

**How can we study identity construction in early English letters:
Combining macro and micro perspectives in linguistic analysis**
Minna Palander-Collin (University of Helsinki)

In language and identity research it is the current view that identities are constructed in interaction as a mixture of individual agency and social structures, but there is no common understanding on the methods of combining micro and macro perspectives for the purposes of linguistic analysis. Moreover, corpus-based methods are infrequently employed and the research focuses overwhelmingly on present-day communities and speakers (e.g. Agha 2007; Auer 2007; Blommaert 2005; Bucholtz and Hall 2005; Omoniyi and White 2006). Consequently, this presentation explores identity construction in a historical period in the interaction of 16th-century personal letters and asks how we could study the language of identity construction in terms of individual agency, on the one hand, and societal structures, on the other, using a combination of corpus-based (clusters, keywords) and context-sensitive discourse analytic methods.

The analysis relies on the sociological domain theory (Layder 1997/2003), which is an integrationist model explaining everyday interaction as a complex web of individual psychobiography, situated activity, social settings and societal resources. Palander-Collin and Nevala (In press) employed this model and showed that these domains can be detected in the way person reference was used in 16th-century letters of Nathaniel Bacon to index salient social hierarchies of the period and to conform to the conventions of the letter genre but also to convey Bacon's personal needs and orientations in a particular situation.

In this presentation, I shall further analyse 16th-century gentlemen's letters sampled from the *Corpus of Early English Correspondence* (CEEC) to explore to what extent it is possible to detect conventional communicative patterns ("macro") and individuality ("micro") in their linguistic identity construction. More specifically, the analysis will be based on self-presentation in son-father letters to attain adequate comparability through similarity in the type of identity being constructed.

“So we've ordered lager shandies, didn't we?”:
The present perfect narrative in varieties of British English and further afield
Jim Walker (Université Lumière Lyon 2)

It is often stated in the literature that the only verb form available in contemporary English for recounting single events in past time is the Preterit. Every single available commercial and academic grammar of the English language supports this view, relegating the Present Perfect to a whole series of other uses, such as experiential past, past events with present repercussions, and so on. However, it is an indubitable fact that in some varieties of English, as is demonstrated by the title of this proposal, taken from a sketch by British comedienne Catherine Tate, the Present Perfect is used, in the absence of past time adverbials, for some forms of narrative. Anecdotally, there does seem to be a strong association with sport, and in particular with football, as the following example demonstrates, where we see a Preterit and a Present Perfect in quick succession:

"We should've gone in 3-0 up at half time, at least. Wonderful football for the first goal, a sizzling shot by Alan, one of his best. But Alan then missed a big chance and their keeper has clattered Kieron outside the area..." (Sir Bobby Robson, at the time manager of Newcastle United)

In this example, taken from a post-match interview, it is initially difficult to see what motivates this switch from Preterit MISSED to Present Perfect HAS CLATTERED. In this paper, I shall demonstrate that this usage is in fact very widespread, in a number of non-standard varieties of English and that it is substantially different from the much documented phenomenon of tense shifting. I shall be reflecting on whether it is simply a non-standard form which has always been latent in various varieties and which has gained a certain 'visibility' as a result of media democratisation, or whether, as some have suggested, English may be travelling down a road towards a narrativisation of the Present Perfect, a path that many other languages have followed before.

However, the success of the foregoing is dependent on resolving the problems inherent in attempting the diachronic study of non-standard forms, and the talk will focus on the methods used (corpora and concordances of 19th century literary texts, corpora of personal correspondence) and whether or not we are doomed to be unable to fully answer the questions posed above.

Developments of Idiosyncratic Constructions

The extremes of insubordination: exclamatory *as if*!

Laurel Brinton (University of British Columbia)

In a draft entry, the OED lists colloquial *as if*!, defining it as ‘expressing dismissive or incredulous contradiction’, used as “a sardonic response to a stated or reported suggestion”. A characteristic example (of the 32 found in COCA) is:

(1) *He thinks you’ll be impressed. As if.* (COCA: FIC 1998)

Exclamatory *as if*! is not treated in the standard grammars of English (but cf. Quirk et al. 1985: 999n, 1417n, 1304n on “exclamation tags”). With the release of the film “Clueless” in 1995, discussions in the popular media attributed the form to “Valley Girl Talk”:

(2) *there’s little trace of Austen’s elegant style in the volleys of “y’know” and “as if!” that pour from the film’s Valley Girl-type characters* (COCA: NEWS 1995)

However, the earliest example given in the OED dates from 1902 and is taken from the American novelist Frank Norris:

(3) *“Maybe he’ll come up and speak to us”. “Oh as if!” contradicted Laura* (Pit i.10)

Monoclausal structures with *as if* receive brief mention in a number of sources (see Huddleston and Pullum 2002:1152; Dancygier and Sweetser 2005:229–30; Lopez-Cuoso and Méndez-Naya 2010:§3.2.1.4; cf. Quirk et al. 1985:842 on *if only* clauses):

(4) *Merriam Webster Dictionary says it’s official, google, the popular computer search engine is now part of the English language as a verb. As if I’m going to \ google you.* (COCA: SPOK 2006)

These exclamatory clauses would appear to arise toward the end of the nineteenth century:

(5) *“To go all that distance for a few papers. As if we did not all know you!”* (1883 Crawford, *Doctor Claudius* [CEN])

Such clauses belong to what Evans (2007: 367) calls “insubordination”, the “*conventionalized main clause use of what, on prima facie grounds, appear to be formally subordinate clauses*”. *As if* monoclauses (and *as if*!) would seem to fall under two of the main functions of insubordinated clauses identified by Evans: modality, e.g. exclamation and evaluation (2007: 403–5) and discourse context/highly presupposed material, e.g. negation (2007:410). Quirk et al. (1985:841) note that insubordinated clauses generally have the force of exclamations, “the omission of the matrix clause being mimetic of speechless amazement”. Evans postulates three stages in the development of insubordinated clauses: (1) a stage at which it is possible to treat the clause as underlyingly subordinate, with the main clause ellipsed, (2) a stage at which the clause has achieved greater semantic specificity and the ellipsed main clause cannot be “restored”, and (3) a stage at which the clause is fully nativized as a main clause.

Using a wide variety of corpora of English, this paper will trace the development of *as if*!, arguing that, consistent with Evans’ hypothesis, it does indeed arise out of monoclausal *as if* (cf. ex. 5 above). First-person evaluative insubordinated *as if* monoclauses (*As if I’m a slave to them!* [COCA]) will prove central to the development, as these are typically negative (cf. Dancygier and Sweetser 2005:229). Also important are other *as if* clauses in which the comparison meaning is “much attenuated” in favor of the truth of the content clause (see Huddleston and Pullum 2002: 1151–2). The loss of the insubordinated clause itself, with only the complementizer (*as if*) being retained as a pragmatic marker, thus represents a further stage in Evans’ sequence.

Evans’ suggestion (2007:374–5) that the development of insubordinated clauses involves initially pragmatization, followed by depragmatization (constructionalization) will come under critical scrutiny in this paper.

A cartoon of an intellectual, a hell of a lot of trouble and a bit of a loose end:
A study of expressive *a(n) N of a(n) N* constructions in the BNC
 Beate Hampe (University of Erfurt)

Inspired by Foolen's (2003) work on "expressive bi-nominals" in present-day English, this contribution presents the results of a corpus study of the 1.624 expressions instantiating the sequence *a(n) N of a(n) (modifier) N* in the all BNC files comprising fictive, biographical and auto-biographical texts (15.926.677 words). Roughly half of those expressions represent unmarked *of*-genitives where the relationships between the two nouns is of a classificatory, quantifying or meronymic kind (a). A substantial subset of the remaining ones present expressive binominals in Foolen's sense, i.e. instances of a special construction illustrated in (b).

(a) *a description of a typical school, a statue of a fiddler, a quarter of an hour, a member of a commission, a part of a deep memory*

(b) *a cartoon of an intellectual, a reed of a woman, a wizard of an engineer, a gem of a nun, a hell of a job, a beast of a complaint, a pearl of a day, a dowager of a car, a trickle of a stream, a hook of a nose*

Foolen characterizes the expressive binominal construction as a double-headed one, both syntactically and semantically, such that the construction achieves a figurative (i.e., more or less clichéd metaphorical or hyperbolic) understanding of, e.g., an intellectual as a cartoon, an engineer as a wizard, or a stream as a trickle. Such an equation (due to shared reference) would, of course, be entirely nonsensical with other *of*-genitives (*a school as a description, *an hour as a quarter, *a commission as a member, *a memory as a part).

Apart from elaborating on Foolen's claims, the collostructional analyses of the corpus data suggest that there are several subgroups of expressive *a(n) N of a(n) N* constructions – allowing for generalizations at several levels of specificity. Some nouns in position N1, such as the hyperbolic *hell* and its variant *heck* are highly productive with a large range of nouns in position N2 forming such sub-constructions as *a hell/heck of a(n) X* and, even more substantively, *a hell/heck of a lot (of X)*. They show signs of semantic bleaching, such that *hell* and *lot* can no longer be equated figuratively and N1 only functions as a kind of expressivity marker.

Apart from that, the corpus results on the sequence *a(n) N of a(n) (modifier) N* do also suggest that nouns which literally serve as vague lexical quantifiers in the *of*-genitive, can also appear as N1 in another subgroup of the expressive *a(n) N of a(n) N* constructions. The noun *bit*, for example, cannot literally act as a quantifier, as the indefinite article excludes mass nouns in position N2. Nevertheless, there is a highly frequent subconstruction *a bit of a(n) X*, which centrally serves to hedge classifications of either persons (c), or experiences (d) that are evaluatively/emotionally highly loaded.

(d) *a bit of a fanatic, a bit of a loner, a bit of a loony, a bit of a womanizer, a bit of a liar, a bit of an expert, a bit of a hero*

(e) *a bit of a shock, a bit of a mess, a bit of a laugh, a bit of a hassle*

The talk will finally compare the results obtained for fiction/biography with the results of work currently in progress, which investigating the sequence *a(n) N of a(n) (modifier) N* in all BNC files containing (i) spontaneously spoken language (9.418.462 words) and (ii) texts from broadsheet newspapers (8.617.465 words).

“I am he as you are he as you are me and we are all together”:

Pronoun Case Variation across Varieties of English

Georg Maier (University of Hamburg)

Over the past decade, the distribution of the different pronominal case forms I – me, she – her, he – him, we – us and they – them in Modern English has been the subject of much linguistic debate (e.g. Angermeyer and Singler 2003; Quinn 2005, 2009). This is true especially for those contexts that permit a choice between the nominative and the objective form, such as ‘it-clefts’ (It is I/ me who does the thinking.), ‘pronoun-NP constructions’ (We/ Us Irish are in it again!), ‘coordinate NPs’ (I/ Me and the girl got married.), ‘it BE sentences’ (Who is it? It is I/ me.) and ‘comparatives’ (There are guys better than he/ him.).

Linguistic theorizing has offered various explanations for this phenomenon: On the one hand, formal approaches to case marking propose that the alternation between nominative and objective pronominal forms is determined exclusively by structural case mechanisms (e.g. Burzio 2000; Chomsky 1993). On the other hand, there are approaches holding that pronominal case selection is best treated as a product of local or positional rules and hypercorrection (e.g. Angermeyer and Singler 2003; Burridge 2004; Hudson 1995; Quirk et. al 1985; Sobin 1997). Still other accounts suggest that pronominal case variation in Modern English is due to a difference between ‘weak’ and ‘strong’ pronouns similar but yet different to those classes in Romance languages (Sweet 1875; Quinn 2005).

Though having different perspectives on the matter, these accounts share common ground in two fundamental aspects: Firstly, there is a general consensus that pronominal case variation is found in well circumscribed contexts. Secondly, most of these studies suffer from a lack of substantial empirical evidence to support their arguments, which is, however, partly due to the limits of existing data bases as most of the contexts which allow for variation of pronoun case forms occur very rarely in corpora (Quinn 2009).

In view of this shortcoming, drawing on a large internet-based corpus compiled for this purpose, this study examines the quantitative and qualitative distribution of pronominal case forms in two rivalling contexts, namely in ‘it-clefts’ and in ‘it BE sentences’, i.e. sentences in which ‘it’ and a form of ‘BE’ are followed by a pronominal subject complement (Quinn 2005), across five major varieties of English.

Results of this study demonstrate that matters of pronominal case selection appear to be much more complicated than has been assumed so far. The data indicate that there are several parameters influencing the choice of a pronominal case form in a certain context, namely person and number, syntactic and pragmatic functions, geographical distribution as well as register (Siemund et al. 2009.).

Internet idioms

QQ More:

An Examination of the Origins and Modern Usage of QQ in Online Forums

Jon Bakos (Oklahoma State University)

Much has been said in recent years about the change in discourse that may have been brought about by the internet and texting. Work by linguists such as Crystal (2006) has confronted the question of whether English is under assault, and whether it is appropriate to use 'Netspeak' in various circumstances. Although 'correctness' has received a great deal of discussion, little has been done to study particular internet terms in their native habitats, like such studies of non-internet words as in the recent biography of 'blotto' in Considine (2009). While many describe the internet as a place where one's expressive options are limited (due to visual context and lag time in conversation), tremendous flexibility and linguistic innovation online are there in the form of many emoticons (:-) , O_o , etc), acronyms (ROFL, OMG), and newly-coined words (wiki, pwnage). This paper will look at one term in particular, QQ, used frequently in online communities for Blizzard Entertainment's "World of Warcraft". QQ is most commonly used as an emoticon to tag someone else's forum posts as excessive whining (The Q's representing 'crying eyes'), but this paper will explore its uses more completely, as well as its evolution over time.

In addition to QQ itself, the presentation will address the difficulties of doing corpus work on an online forum. When a researcher is limited to hunting for concordances using the forum's engine, he or she must devise searches to maximize retrievable tokens. Even if one is limited to the first 100 results, it is possible to push the envelope through creative searching and have access to more data. The presentation will also discuss how best to narrow one's focus when looking at a large online speech community, so as to study the most current and representative usages of a term. The most general findings of the study are that QQ's usage is impressively diverse. Due to its construction out of letters rather than symbols, QQ can also be realized in nominal, verbal, or adjectival varieties. While different forms all share the same meaning, there is evidence that verbal varieties may be temporary (One can start or stop 'QQing' or complaining), but that nominative forms are more permanent (a 'QQer,' is branded for life).

Eco-fundamentalist, cougarlicious and co.:
The NeoCrawler as a new tool to identify and observe neologisms on in the Internet
Daphné Kerremans, Susanne Stegmayr and Hans-Jörg Schmid
(Ludwig-Maximilians University Munich)

Language change has been one of the most extensively studied areas in linguistics and a magnet to scholars hoping to discover the invisible mechanisms underlying language. Despite being surrounded by language, linguists have always faced the problem of how to capture and preserve a sufficient, authentic language sample for detailed study. It is merely a few decades ago that the development of corpora boosted up linguistic methods and (r)evolutionized language studies. Since the commercialization of the Internet, a new wave of possibilities and opportunities has arisen, but nevertheless researchers still often feel like looking for the needle in the haystack: an enormous, gigantic collection of real language scattered among an even bigger jumble of photos, videos, computer code, adds etc. Therefore, researchers developed web crawlers that would scan the Web for relevant linguistic data and make it useful for scientific investigations (cf. Renouf, Kehoe and Banerjee 2005, Kilgariff 2003).

Our interest in lexical innovation, i.e. monitoring the behavior of neologisms, has forced us to develop a web crawler, the NeoCrawler, which is tailored to our specific needs. The NeoCrawler exists of two modules. The Discoverer identifies potential neologisms by extracting letter combinations which have not been recognized as words by Google N-Grams (cf. Evert 2010). The second module, the Observer, performs weekly crawls for each of the neologisms that are stored in the database. The results are extracted and shown in a 500-character cotext, together with the original page. The Observer is furthermore equipped with several tools such as a tokenizer and parser to facilitate further analysis. In addition, a two-tiered classification scheme, different from Biber's 2007 proposal, has been developed to study the social, semantic, stylistic and pragmatic behavior of new lexical units that have recently entered the language. Our paper will present the NeoCrawler and its functions as well as illustrate its potential in 21st-century linguistics.

Development of SMS language from 2000 to 2010: A comparison of two corpora

Úrsula Kirsten Torrado (University of Vigo)

SMS language is regarded as a ‘new’ communication system (cf. Shortis 2007; Crystal 2009; Markus 2010) characterized by new relationships that native speakers establish between English spelling and pronunciation by using different respelling devices (cf. Thurlow 2003; López Rúa 2007; Crystal 2009). This paper is an attempt to contribute to this recent area of study by analyzing the development of SMS language over the last 10 years. Recent findings suggest that even though SMS language might have emerged out of the need of speed and brevity (Thurlow 2003: 4) —every SMS has a limited amount of characters—, it seems to have evolved into a fashionable and stylish way of writing where shortened versions of the words are not always the aim of the respelling. For the purpose of a diachronic analysis a free online British SMS corpus available at netting-it.com containing 201 text messages compiled in 2000 has been used in comparison with my own data obtained in May 2010 by means of questionnaires carried out in a London high school. Thus, it was possible to analyze the differences in the usage of SMS language during the last decade. The research proves that the most significant changes are the usage of ‘stylish talk’, a new device which consists in lengthening words to emphasize accent, slang, and attitude. This contrasts with the general belief that words are merely shortened in SMS language. Moreover, the usage of slang and agrammatical expressions seems to be frequent devices as well.

Corpus Studies

Flexibility and Application of the Bounded Virtual Corpus

Garrison Bickerstaff (University of Georgia)

The purpose of this paper is to highlight a new methodology in corpus linguistics referred to as the Bounded Virtual Corpus (BVC). The American BVC uses 25 full-text U.S. newspapers in the LexisNexis Academic database to research the recent careers of forms in American English.

For comparison, a British BVC is constructed by a separate collection of 5 full-text British newspapers in the same database. The titles in each BVC cover a span of ten years—1998 to 2007. The construction of the BVCs can be duplicated and, more importantly, can be altered to fit the exact needs of another researcher. The British and American BVCs can answer questions about the recent careers of forms in British and American English, such as whether a particular form is current in both American and British English.

The construction of the American BVC allows identification of regional patterns of word use in American English as 5 full-text American newspaper titles are included in the American BVC from each of 5 regions—northeast, southeast, midwest, west and coastal west.

This paper will highlight how the methodology of the BVC can assist three research applications. First, the recent career of the British construction *went missing* in American English is examined. The form *went missing* appears 46 times in 1998 in the American BVC; in a 2007 search, the same form appears 594 times. Such useful information is accessible because of the extremely large amounts of text in the BVC (an estimated 5 billion words in the American BVC and an estimated 1 billion words in the British BVC). Also, the articles in which search forms appear are readily accessible in full-text format.

Through evidence from the BVC, the paper will also examine the career of the form *carb* that is related to the low-carb diet. The BVC will show evidence that relates to the career of the form *carb* in American English before, during, and after (until the end of 2007) the extreme popularity of the low-carb diet. For a view of the bigger picture, the paper will also show evidence from the BVC of some of the collocating habits of the form *diet* away from the specialized language of the low-carb diet.

Finally, the paper will examine how the BVC can produce a broad range of high and low frequency collocates of the form *nickel*. The paper will show with the form *nickel* that the methodology of the BVC makes it a useful tool for lexicography.

Tools for comparing corpora

Terttu Nevalainen, Tanja Säily, Harri Siirtola (University of Helsinki)

The number of electronic corpora has increased substantially over the last decades. The Brown family of corpora, for example, has gained new members which both update the original 1960s versions and extend them back in time to the beginning of the 20th century. Using the same sampling frame has the aim of optimizing genre continuity. Although not unproblematic or even obtainable in long diachrony, access to comparable data is a prerequisite for the study of language variation and change over time.

Earlier research shows that, for a variety of reasons, genre continuity is not easy to achieve in practice, and even where it is, the ‘same’ genres can vary considerably in their linguistic make-up (e.g. Biber & Finegan 1997). The aim of our paper is not to discuss the reasons for this variation but rather to provide tools for comparing corpora and hence for spotting such differences.

Some tools that may be of use are visualization techniques based on public domain resources such as Many Eyes (IBM Visual Communication Lab 2010). They have a certain heuristic value but rarely provide enough information for comparing corpus differences and similarities in sufficient detail. While certainly useful, tools such as Wmatrix (Rayson 2009) that have been developed for corpus-linguistic purposes do not fully exploit the possibilities offered by information visualization. More flexible and interactive tools are being developed in the multidisciplinary DAMMOC project (Data Mining Tools for Changing Modalities of Communication). We will introduce one such tool, Text Variation Explorer, which allows the comparison of data sets of varying sizes using a number of parameters that have been found to be helpful diagnostics in distinguishing genre variation (cf. Biber & Finegan 1997). These parameters include type/token ratio (TTR), word and sentence length, and hapax legomena.

Figure 1 shows the user interface of Text Variation Explorer, which allows rapid exploration of the effects of parameter window size on measurement variation by manipulating the window size slider and observing the changes in the visualization. The text and the visualization are linked together in such a way that a selection in one of them is highlighted (“brushed”) in the other. In addition, Text Variation Explorer can zoom into parts of text, ensuring that each text fragment will be represented at least by one picture element in the visualization. The tool can be used, e.g., for comparing the TTR and word-length profiles of two data sets, identifying the peaks and valleys as well as the range of variation in each at a suitable window size. This is a much more comprehensive (and yet easy-to-grasp) comparison than that between single measurements such as the average TTR of a fixed-size sample, and thus better suited to identifying genre variation.

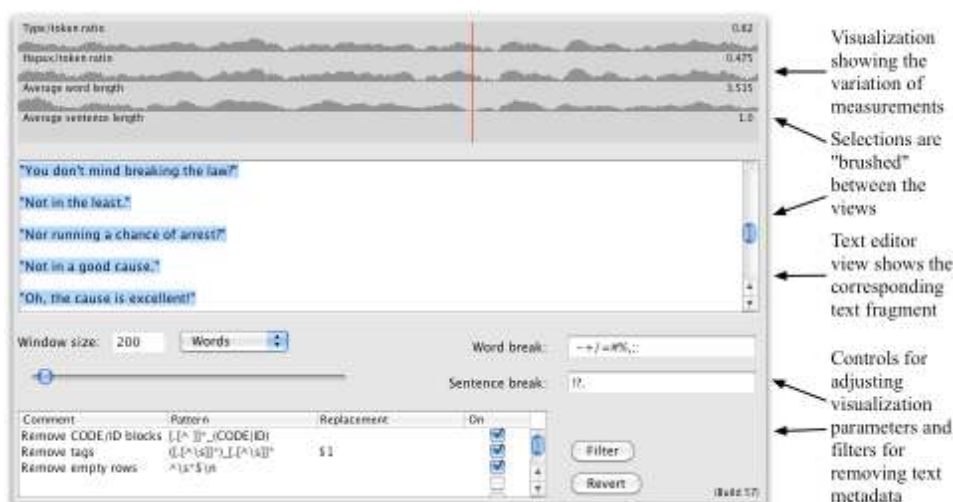


Figure 1. Text Variation Explorer.

The adjusted frequency list:
A new method to extract cluster-sensitive frequency lists from corpora
Matthew B. O'Donnell (University of Michigan, Ann Arbor)

The suggestion that language learners acquire and make use of multi-word chunks without either breaking them apart or building them up from individual words is well established in psycholinguistic (Ellis 2003) and corpus linguistic (Sinclair 1991) circles. It is now even discussed in the popular press (Zimmer 2010). Frequency lists of items of various lengths are important in both computational and applied linguistics. O'Keefe et al. highlight the fact that 'many chunks are as frequent as or more frequent than the single-word items which appear in the core vocabulary' (2006: 46). They are also valuable for measuring the idiomatic/formulaic nature of text (Erman & Warren 2000). However, many of our computational tools and methods still focus on individual words as the foundational units of analysis. The method proposed here is designed to address this issue.

A common methodological step in a corpus-linguistic analysis is the extraction of frequency lists of various size chunks (called clusters, lexical bundles or n-grams). Most software packages facilitate the creation of such lists, making it possible to compare units of differing lengths. However, each size unit is counted on its own terms. For example, imagine a text in which every instance of *know* is preceded by *you*. In a combined 1- and 2- gram list *know* and *you know* would have the same frequency even though *know* does not have any independent occurrences to be counted. The issue also applies to larger units where collecting 3-, 4- and 5-grams together will result in very similar and often identical counts for *at the end*, *the end of*, *at the end of*, *the end of the*, *at the end of the*, *the end of the day* and so on.

The index-based method of constructing a frequency list proposed here adjusts the frequency of items of various lengths if they are part of a larger unit that occurs at or above a given frequency or statistical threshold. That is, if *you know* occurs 15 times in a corpus and *know* 20 times, then the frequency of *know* will be adjusted from 20 down to 5. The method outlined is 'cluster sensitive' because it boosts the rank of larger word sequences and builds on the notion that if such chunks are single choice items for speakers they should be counted as single items and internal constituents left uncounted. Using sections of the BNCBaby corpus, a number of comparisons of unadjusted (standard) and adjusted frequency 1- to 5-gram lists will be presented. For instance, the top 10 items in a standard combined 1- to 5-gram frequency lists of the BNCBaby Demographic section are: *i*, *you*, *the*, *it*, *and*, *a*, *to*, *that*, *yeah*, *oh*. Applying the adjusted frequency method produces a cluster-sensitive list with the following top 10 items: *i don't know*, *and*, *the*, *do you want*, *one two three*, *i don't think*, *of*, *in*, *two three four*, *a*.

The simple method presented here, along with other more complex techniques that have been recently proposed (Gries & Mukherjee, 2010), demonstrates how corpus analysis continues to validate the importance of chunking in the investigation and description of language.

Case Studies

Call so and so and tell him such and such:

A corpus-based study of suspensive reference in contemporary English

Lieven Vandelanotte (University of Namur and University of Leuven)

The items *so and so* and *such and such* along with their hyphenated and (at least for *soandso*) unhyphenated spelling variants (henceforth collectively referred to as *S and S*) can naturally be grouped together as substitutes standing in for more specific expressions “not exactly remembered or not requiring to be explicitly stated” (*OED* sv *so-and-so*); a euphemistic avoidance strategy has also been invoked to account for some uses of *so and so*. This paper argues that *S and S* are best analysed as suspending reference, and reports on a study of their grammar, meaning and use in the Corpus of Contemporary American English (COCA, Davies 2008–).

The association of *S and S* with contexts of direct and indirect speech constructions was already noted in early studies of their German counterparts (Harweg 1970, 1972), and in this context it was noted that *S and S* do not only disallow a *de dicto* reading, as they cannot be understood as the represented speaker’s original designation, but also a *de re* reading (von Roncador 1988: 108), as they barely refer at all. From a theoretical standpoint, *S and S* can be viewed as ‘suspensives’, a term reserved in the so-called Pronominal Approach (van den Eynde et al. 2002) for words like *who*, *where*, *when*, the difference being that the latter carry clear higher-order type specifications (e.g. ‘person’, ‘time’, ‘place’) which *S and S* lack.

Descriptively, analysis of the COCA corpus data allows a more fine-grained understanding which moves beyond the typical usage and patternings, which include for *so and so* (suspended) reference to people (*Mr So and So*), including the polite avoidance of taboo words referring to people (*lucky old so and so*) and for *such and such* the predeterminer usage in *such and such a day*. For instance, while *S and S* do strongly favour reported speech contexts, with a preponderance of direct over indirect speech, both forms also occur in non-reported contexts, for instance where a character is evoked in a hypothetical scene (*...hours in a traffic jam to get to a cold damp cottage where dear Mrs So and So from the village has not been able to switch on the heating*). Furthermore, it turns out that *S and S* show areas of overlap, where one occurs in a pattern more typically associated with the other, as in *so and so ship has arrived* (pronominal use less common for *so and so*) or *the university of such and such* (proper name use less common for *such and such*). Puzzles in the grammatical analysis of *S and S* include determinerless ‘predeterminer’ uses as in *such and such song* and extended uses of *such and such* where, unlike in core cases, replacing the expression by mere *such* results in ungrammaticality (*take a job at such and such*). While *S and S* remain highly unspecific substitutes, a semantico-pragmatic analysis of what the expression contextually stands in for, which is co-determined but not fully predicted by the different syntactic functions involved, helps flesh out the differentiation between the two expressions studied.

Against the backdrop of a cognitive-functional model of the English noun phrase (e.g. Breban 2010: 11–39), this paper sets out to inventorize and explicate the different uses of *S and S* in contemporary English across the five register types represented in COCA, where the main dividing line is between spoken data, where *S and S* (and particularly *so and so*) is most frequent, and the fiction, newspaper, magazine and academic sections of the corpus.

A new perspective on antonymy: What corpus-based studies of nouns can tell us
Gregory Garretson (Uppsala University)

Antonymy, or semantic opposition, has been a topic of interest to linguists for centuries. However, most investigations of antonymy have two things in common: first, they are based on introspective analysis rather than on corpus data, and second, they are based solely on adjectives (see, e.g., Cruse 1986, Jones 2002). This study of antonymy breaks with tradition in two ways, by using a data-driven, computational, corpus-based methodology, and by focusing on other parts of speech, specifically nouns.

The study addresses three research questions by making use of new corpus-linguistic techniques and two rich sources of data: the British National Corpus (Burnard 2007), a 100-million-word, balanced corpus of modern English, and WordNet (Fellbaum 1998), a database of English lexical relations describing the relations among over 150,000 words.

The first research question is whether nouns make —good antonyms in the way that adjectives are said to do. In other words, is the relation between *day* and *night* indeed similar in nature to the relation between *early* and *late*? As will be seen, it turns out that nominal antonymy is non-prototypical to a great degree.

The second research question asks whether it is possible to design an algorithm that successfully matches nouns to their antonyms using corpus-derived information on the words' *semantic preference relations*—that is, their tendency to co-occur with words from particular lexical fields (Sinclair 2004). For example, the adjective *large* has a tendency to co-occur with words such as *number*, *scale*, *part*, *amounts*, *quantity*, etc., all of which have to do with —quantities and sizes (Stubbs 2002:65). The source of the semantic preference data in this study is a computer program I have developed that profiles words based on their syntagmatic relations—such as collocation, colligation, and of course, semantic preference—in large corpora (cf. Sinclair 2004). It is found that it is indeed possible to design such an algorithm, provided that the size of the candidate pool is under a certain threshold.

The third research question takes the analysis even further, asking whether semantic preference information may also be used to detect asymmetries *within* pairs of antonyms. Indeed, it is found that many pairs of antonyms exhibit a kind of asymmetry that differs from the types of markedness previously described, with one word being what I term the —strong member (e.g., *day*, *front*, *war*) and the other the —weak member (e.g., *night*, *back*, *peace*). A discourse-based explanation is offered for this systematic asymmetry between nominal antonyms.

An advantage of studying antonymy in nouns rather than in adjectives is that nouns provide a great wealth of non-prototypical cases of antonymy. In fact, the investigations described here lead me to propose a refined model of antonymy, in which prototypical antonymy is achieved only if both a semantic constraint and a lexical constraint are satisfied (cf. Justeson & Katz 1992). The model also has the advantage that it allows us to identify and explain various types of non-prototypical or —aberrant antonymy; these are, in fact, shown to account for a majority of pairs of nominal antonyms. As a result of this investigation, we are provided with a more complex and more complete picture of antonymic relations in English.

Corpus linguistics with Google?

Stefan Diemer (Saarland University)

This paper aims at revisiting the issue whether it is possible to use commercial web search tools such as the Google interface for meaningful corpus research, given recent advances in search technology. It is argued that with the proper methodology these web tools can and should be used for corpus research, since they provide considerable advantages in comparison with both closed corpora and web-based linguistic search tools.

Previous investigations of this issue have been mostly negative. Fletcher (in Hundt et al. 2007) is just one of many citing procedural problems, a criticism that is illustrated by several web-based sample studies, (Rosenbach, *ibid.*). Other researchers advocate the use of specific linguistic search tools like WebCorp, while admitting that these are less efficient than their commercially developed counterparts. More general issues are the question of representativeness and reliability of results (Leech, *ibid.*). It is not surprising, therefore, that there are few previous studies using a Google-based corpus (among them Lindvall 2004).

However, this mostly negative verdict may have to be reconsidered in view of powerful tools recently developed by Google, specifically “Trends” and “Insights for Search”. Both use Google’s infrastructure for web search, but provide considerably more data than the standard interface. “Google Trends” gives researchers the option of instantly examining search volumes for specific search terms or phrases, either diachronically or synchronically. The data is provided in numerical, graphical and normalized formats and can be adjusted for region, country, city or language. It is possible to plot developments in lexical or phrase search use and to compare search terms, using different variables such as geographical or diachronic distribution. “Insights for Search” uses the same data as “Trends”, but is aimed at researchers (or advertisers) and includes several additional features, including geographic “heat maps” that illustrate the dissemination of search terms, and a forecast option that calculates the potential future use of search terms. The inclusion of blogs and other online text types in the corpora allows the examination not only of past, but also of ongoing language innovations and changes.

Following Diemer (2008), it will be argued on the basis of several examples that these tools can enhance both qualitative and quantitative corpus research, particularly in the areas of lexis, phraseology, sociolinguistics and cognitive semantics. Possible applications in lexis and phraseology include the quantitative and qualitative documentation of lexical or phraseological innovation in combination with a geographical component and the investigation of word and spelling changes and the use of non-standard features. The paper will also present results from a recent study (Diemer 2010) illustrating an increased web-based use of prefixed verbs in English, possibly reversing a 1000-year decline. In sociolinguistic research, the tools allow the precise description of sociolects based on real-time data, while in cognitive semantics this approach can pinpoint connotations much more convincingly than existing corpus-based methods.

The paper further discusses a suitable methodology for using these tools, taking into account limitations such as corpus size, the lack of tagging and in particular the issue of representativeness.

The constituency of hyperlinks in a hypertext corpus

Michael Yoshitaka Erlewine (Massachusetts Institute of Technology)

While much recent work has advocated the use of Internet corpora for traditional corpus linguistics and some have looked at the effects of hyperlinks on narrative structure (Bexten, 2006), no previous work has used one of the greatest characteristics of hypermedia—inline hyperlinks—as a tool in the generative study of syntax. In this talk I present a novel methodology for using hyperlink corpora as a test for syntactic constituency.

Consider the following sentence (with **bold** text indicating hyperlinked text) from the website MetaFilter (<http://metafilter.com>):

(1) October's **focus on breast cancer** is a **curvy double-edged sword** and those **in the fight agree**.

Intuitively, every hypertext link represents a referent and is meant to give the user some semantically appropriate information on what to expect when clicked. For this reason, we may expect the link text to be a linguistic unit—corresponding to the notion of a *constituent* in generative syntax—in its host sentence.

Given this hypothesis, we note that the final string of link text, “in the fight agree,” does not itself form a natural linguistic unit, as “in the fight” is part of the noun phrase “those in the fight.” However, this potentially non-constituent hyperlink is an illusion of the bold-faced marking in (1) (as is on the MetaFilter site). In reality, “in the fight agree” is actually made up of two links:

(1) ... and those [**in the fight**] [**agree**] .

Examples such as this show how hyperlinks can serve as a window into the mental representations of linguistic structure which underlie sentences.

In this study a hyperlink corpus of over 4.8 million words was constructed from the MetaFilter website, a link-rich community weblog. Host sentences of links were parsed using the open-source Stanford Parser (Klein, 2003), trained on the Penn Treebank (Marcus, 1993). Strings corresponding to hyperlinks were then checked for constituency according to these parses. Manual supervision was used to refine these results in order to combat common sources of error.

We conclude that hyperlinks correspond very strongly to the syntactic notion of constituency, establishing this novel hypertext methodology as a potential constituency test. Examples of hyperlinks which would not be considered constituents through more traditional constituency tests are then discussed, as they may provide arguments both for and against differing notions of constituency itself. Such “non-constituent links” are compared to notions of constituency in various traditions in generative syntax, including that of Combinatory Categorical Grammar and the Minimalist Program.

Posters

Death-related intensifiers: Grammaticalization and related phenomena in the development of the intensifier *deadly*

Zeltia Blanco-Suárez (University of Santiago de Compostela)

From an anthropological point of view, death is a matter which pervades our daily lives, and is thus an issue of the most genuine concern for all cultures and societies worldwide. With such an impact on our routines it comes as no surprise that it can be exploited not only as a linguistic taboo, but also as a source of intensification in language, perhaps even crosslinguistically.

This paper will in broad terms tackle the issue of intensification in language, and will concentrate on the intensifier strategy in particular (cf. Bolinger 1972). More specifically, it will focus on one of the death-related intensifiers in English, to wit *deadly*, by tracing its evolution over time, and by additionally showing how the counterparts of *deadly* in other languages might well undergo similar evolutionary paths.

My preliminary data seem to suggest that *deadly* has undergone a grammaticalization process in the course of time. Thus, it was originally an adverb and adjective applying to entities which could entail physical or spiritual death, as in (1) and (2) below:

(1) *Bei seye þat wee synne **dedly** [F mortelement] in schauynge oure berdes.* (c. 1425, *MED*, s.v. *dedli* adv. 2).

(2) ***Dedli** drynke, ȝif Bei taken it..anoieþ hem not.* (c. 1380, *OED*, s.v. *deadly* a. 4b). Then *deadly* gradually developed more subjective meanings, relating to qualities typically associated with death, such as paleness, darkness, or silence (cf. (3) and (4)). It was precisely these subjective meanings which provided the bridging context for *deadly* to be used as an intensifier, as in examples (5) and (6):

(3) *There was such a **deadlie** silence in the porte.* (1600, *OED*, s.v. *deadly* a. 7c).

(4) *Custaunce, with a **dedly** pale face..toward hir ship she wente.* (c. 1390, *MED*, s.v. *dedli* adv. 3).

(5) *I þat es sa **dedli** dill* [‘stupid’]. (c. 1300, *OED*, s.v. *deadly* adv. 4).

(6) *A **deadly** drinker he is, and grown exceedingly fat* (1660, *OED*, s.v. *deadly* a. 8a).

Instances (1) to (6) above, therefore, clearly evince the gradual subjectification (cf. Traugott 1995, Traugott and Dasher 2002, and Athanasiadou 2007, among others) and grammaticalization (cf. Heine 2003, and Hopper and Traugott 2003, among others) of *deadly* through time, thus tallying with other grammaticalization clines observed in the literature: descriptive adjective > affective adjective > intensifier (Adamson 2000); qualitative adverb > booster (Peters 1994); ‘modal-to-intensifier shift’ (Partington 1993).

In other European languages such as French, German, Italian, Portuguese, and Spanish similar collocations to those found for *deadly* in English are attested. Can we then safely assume that *deadly* in these languages has undergone a parallel development, and that equivalents of *deadly* will also be found in other non-European languages, with similar results, on account of the momentousness which death has on our routines? Is this transferable to other death-related intensifiers? These issues will be delved into in the present study.

Data for the present paper will be retrieved from a variety of diachronic and synchronic corpora, including the *Helsinki Corpus of English Texts*, the extended version of the *Corpus of Late Modern English Texts*, and the *British National Corpus*. The online editions of the *Middle English Dictionary* (*MED*) and the *Oxford English Dictionary* (*OED*) will be used as additional sources of evidence. As for the examples from languages other than English, these will be taken from monolingual dictionaries of the respective language.

Shall we *start* or ... *commence*?: Stylistic aspects of near-synonymous verb use
Daniele Franceschi (University of Pisa)

The aim of this paper is to analyse the different stylistic effects produced by the use of three near-synonymous verbs in context (Murphy 2003), i.e. *start*, *begin* and *commence*, as part of a wider project investigating syntactic and semantic-pragmatic variation between Latin-based lexical items and their Anglo-Saxon correspondents (lexical pairs). The study is corpus-based (BNC) and restricted to British fiction and non-fiction texts produced between 1985 and 1993. It follows a sign-oriented approach based on the concept of invariant meaning (Tobin 1989; Contini-Morava & Tobin 2000), markedness (Merlini Barbaresi 1988b, 2003b) and complexity (Bertuccelli Papi 2003; Bertuccelli Papi & Lenci 2007).

The analysis shows that *start*, *begin* and *commence* share some basic semantic features, but appear in different lexico-syntactic alternations and constructions. The frequency of occurrence of *start* and *begin* is generally higher if compared to that of *commence* which prevails in informative text types and/or in stylistically marked contexts, i.e. *start* and *begin* are broader in the scope of their application to the field of reference. When the verbs in question are semantically indistinguishable and interchangeable in terms of plausible formal realisations, the writer makes a deliberate stylist choice to serve a specific purpose. The effects produced are not just the reflection of differences in register (formal vs. informal) or between genres (scientific vs. literary): they are also motivated by the writer's intention to express his or her own subjectivity more directly.

The prototypical behaviour of *start*, *begin* and *commence* can be subject to a process of stylistic specialisation aimed at achieving a pragmatic goal, e.g. a sense of solemnity, seriousness, irony etc.

Natural selection in self-organizing morphological systems

Mark Lindsay and Mark Aronoff (Stony Brook University)

In recent years, it has been shown that certain subsystems of human languages can be profitably investigated in terms of self-organizing or emergent systems (Baayen (1993), Anshen and Aronoff (1999), Goldsmith (2001), Albright (2002), Goldsmith (2006), Plag and Baayen (2009)). Using new methods made available by the Internet and modern databases, we will show that the birth of productive affixes from borrowed vocabulary can be treated as an emergent system, where affixes survive or perish due to circumstances of environment and competition. We present two types of evidence for this position.

The suffixes *-ity* and *-ment* began as borrowings from French in the 13th century; the rate of borrowings of words containing these suffixes peaked in the 17th century, at which point both began a steady decline. The number of derived forms also increased as borrowings increased, but in the 17th century, these two suffixes diverged: *-ity* flourished as a derivational suffix, while *-ment* began a sharp decline that has resulted in nearly zero derived forms today. Why were the two fates so different? By their very nature, productive affixes exist in morphological ecosystems, dependent on new words as a source of productivity. While *-ment* is dependent on verbs to create new words, *-ity* is dependent on adjectives. If we look at the percentage of new verbs and new adjectives over the last 800 years, we see a sharp decline in the number of new verbs beginning in the 17th century, while adjectives steadily increased. This decline in new verbs directly coincides with the decline of *-ment*. Thus, *-ment* died out because it presumably did not have enough new verbs to maintain a viable productive ecosystem.

Another type of evidence comes from the study of the rival suffix pair *-ic* and *-ical*. We call the pair ‘rivals’ because they are in synonymous. Why is synonymy tolerated in language? In most *-ic/-ical* pairs, one form is strongly preferred over the other: *electronic* is much more common than *electronical*, while *surgical* is much more common than *surgic*. What is novel about this research is the way in which we measure productivity: the number of Estimated Total Matches (Google hits) for each word ending in each suffix. For every word, we do a Google search on the exact term and list the number of hits; we then look for numerical patterns in these numbers to determine productivity.

Using Webster’s 2nd International dictionary, we identified a total of 11966 stems of English words ending in either *-ic* or *-ical*. We then queried Google for words ending in both *-ic* and *-ical* in each of the 11966 stems. For some stems, a Google search will record hits for both *-ic* and *-ical* words (e.g. *historic* and *historical*), while for some, only one form has any hits. We then determined for each pair whether *-ic* or *-ical* had more hits; this word was considered the ‘winner’. Ninety percent of all comparisons yielded a winner by a margin of at least one order of magnitude. Using this comparison, we identified 10613 *-ic* winners vs. 1353 *-ical* winners with a ratio of 7.84 in favor of *-ic*. This demonstrates conclusively that, overall, *-ic* is more productive than *-ical*.

Finer-grained analysis, however, reveals that *-ical* is potentiating (Williams 1981) within a certain domain. To investigate this, we first sorted all the stems in our database into left-to-right alphabetical neighborhoods of one to five letters, e.g. all stems ending in *-t-* (there are 4166 of these). When we sort the words in this way, we find that the only set

of words ending in *-ical* with a neighborhood over 100 in size is *-olog(ical)*; for this subset only, *-ical* is the winner over *-ic* (e.g. *psychological* over *psychologic*) by a ratio of 8.30, almost the exact reverse of the ratio of the full set. In other words, using this Google search technique, we find that, although overall *-ic* is more productive than *-ical*, the reverse is true for words ending in *-olog(ical)*.

Although *-olog(ical)* forms a large subset (475 members), no other large sets in the system have resisted the overall trend favoring *-ic*. However, the *-olog(ical)* subset is unique in another way: the string ‘olog’ has a strikingly low number of competitors. For example, there are 79 stems ending in *-rist-*, but this number jumps to 4166 stem ending in *-t-*. Thus, *-rist-* makes up only 1.9% of stems ending in *-t-*, leaving many similar competitors. We found that, on average, a neighborhood of size 2 (e.g. *-st-*) only accounts for 27.84% of the words in the corresponding neighborhood of size 1 (e.g. *-t-*), while a neighborhood of size 4 makes up just 10.47% of its size-2 set. However, even at size 4, the *-olog-* subset still makes up 66.62% of its size 3 (*-g-*) set. This means that 66% of words ending in *-gic(al)* also end in *-ologic(al)*, which is the strongest of all sets by a wide margin (the next strongest is *-graph-* at 34%). Thus, *-olog(ical)* is a subsystem that is not only sufficiently large, but also has distinctly few competitors, leaving it uniquely suited to sustain itself in spite of patterning inversely to all other *-ic/ical* pairs.

We know from our historical study of *-ity* that borrowed suffixes in English form emergent self-organizing systems. If words ending in *-ic* or *-ical* formed a simple emergent system, we might predict, based on the overall preponderance of *-ic* words, that this rival would eventually win out and that *-ical* would lose. Instead, we see that a strong regularity, even one that is the reverse of the normal pattern, can develop in a subset if the subset stands out. We expect to explore other such subsets in further studies. For example, we know that *-ity* and *-ness* are rivals and we have a variety of types of evidence showing that *-ness* is more productive, with some evidence that *-ity* is potentiating in a subset of words (words ending in *-ible* and *-able*). We will use the same Google search techniques to explore the set of words ending in either *-ness* or *-ity*, to see whether there are structural commonalities with the set of words that we have explored here.

Overall, and somewhat surprisingly, English derivational morphology, especially when it involves the emergence of productive affixes from sets of borrowed words (in which English is especially rich), is a fertile proving ground for the study of self-organizing systems in languages, in part because of the databases that electronic resources provide.

Imperfectives in Singapore English: new evidence for ethnic varieties?

Jakob R. E. Leimgruber (University of Freiburg)

Lavanya Sankaran (Queen Mary, University of London)

Studies of Singapore English (SgE) showing differences in ethnic varieties have been largely limited to phonological variables and to identification tasks (Deterding and Poedjosoedarmo 2000, Lim 2000, Suzanna and Brown 2000, Tan 2005, Deterding 2007). Grammatical variables have not generally been found to vary along ethnic lines. On the other hand, general works in World English studies (e.g. Mesthrie 2008) agree that there are several grammatical differences between SgE and Indian English (IndE), for instance in the use of imperfectives, which are extended to cover statives and non-delimited habituals in IndE but not in SgE (Sharma 2009). In these comparative studies, however, SgE is always taken as a whole, a whole which, as shown in the literature (Ho and Platt 1993 *inter alia*), is actually quite heterogeneous.

A recent pilot test on SgE speakers from the three major ethnic groups (Chinese, Malay, Indian) shows that there may be differences in how individual ethnic groups mark non-delimited habituals in their sub-variety. It is indicated that Indian (Tamil) speakers use the imperfective more for non-delimited habituals than the other groups. If confirmed, this trend would be a tangible grammatical difference in ethnic sub-varieties of SgE, which have until now been discriminated by and large only phonetically. Furthermore, such a trend would indicate that Tamil speakers of SgE exhibit some similarities to speakers of Indian English in how they mark non-delimited habituals. We argue that the root for this would lie in Tamil, where non-delimited habituals, like statives, are marked by the imperfective.

Diachronic corpus linguistics: overcoming the limitations of automatic analysis

Nadja Nesselhauf (University of Heidelberg)

It is indisputable that the possibility of automatic analysis of authentic language data is one of the great advantages of modern corpus linguistics. Naturally this is also true for the corpus-linguistic investigation of many areas of language change. However, the possibility of automatic analysis also bears some dangers. Not only is there the danger of largely restricting corpus-linguistic analyses to results that can be obtained (semi-) automatically but, more fundamentally, there is the danger of tailoring research questions to match the technical possibilities.

In this poster, the limitations of such a methodology-driven approach will be made evident for an important area of semantico-syntactic change in recent centuries: the changes in the way 'future' is referred to in English. It will be demonstrated how a scientifically more sound approach of the methods being determined by the research question may lead to a more accurate and more comprehensive account of linguistic changes. Such an approach therefore constitutes a much better foundation for uncovering mechanisms of language change.

In the light of the above, it is unsurprising that the diachronic investigation of the expression of 'future' in English has so far largely been restricted to *will*, *shall*, *'ll* and *BE going to*, where future reference is tied to strings that are easily retrievable automatically (e.g. Berglund 2005). In addition to that, in particular studies of (or including) *will* and *shall* are often merely based on total counts of the expressions and do not disambiguate the different senses, so that instances that cannot be considered future time expressions (such as *shall* in the sense of 'obligation' or *will* in the sense of 'willingness') are also included (e.g. Kytö 1991). Laudable (but rare) exceptions to this trend are Gotti (2003), where on the basis of manual disambiguation the non-future uses of *shall* and *will* are excluded (but the analysis is restricted to these two expressions), and Smith (2005), where the diachronic development of several future time expressions is investigated at once (but the senses are not disambiguated).

The study presented here tries to overcome both these restrictions simultaneously. All major ways of referring to the future are considered no matter whether they are easily retrievable automatically or not and a detailed manual analysis is performed on the data so that only those senses of the relevant expressions which are actually future-referring are considered. In addition to *will*, *shall*, *'ll* and *BE going to*, the future time expressions investigated are the simple present, the present progressive with future time reference, and *BE to*. The time span considered is 1650-1999; the corpus on which the analysis is based is ARCHER (*A Representative Corpus of Historical English Registers*).

Some preliminary findings from this analysis indicate that several of the earlier results which stem from studies with the methodological restrictions described above indeed to have to be revised (cf. also Nesselhauf 2010). The investigation reveals, for instance, that in recent centuries a rise of the present progressive with future time reference has taken place, which has hardly been noticed so far, although it appears to be of about the same extent as the widely discussed rise of *BE going to*. Furthermore, the wide-spread assumption of the "rise of *will*" in the past few centuries (e.g. Nurmi 2003: 105; Görlach 1999: 84) seems to be a consequence of the restriction of many studies to *shall* and *will* and a subsequent interpretation of the results only in relation to each other. The present more comprehensive study much rather indicates that *will* has roughly remained stable, and that the decline of *shall* has largely been made up for by the rise of other expressions. It is thus demonstrated how diachronic corpus linguistics can profit from overcoming the limitations of automatic analysis, as these limitations are self-imposed by researchers and not inherent in the corpus approach.

Introducing visuals to historical pragmatics: Book history and multimodality

Carla Suhr (University of Helsinki)

This paper will show how pragmatic study of historical materials can be augmented by methods from book history and multimodal studies, that is, by introducing visual aspects of a text into its pragmatic analysis. It has long been recognized that the study of historical texts requires knowledge of their social and cultural settings, but so far historical linguists have paid little attention to the actual physical text and its production process. The reality is that, like medieval manuscripts, an early printed text was a collaborative effort: various typographical features and illustrations were decided on in the printing house by someone other than the author (see e.g. McKerrow 1928). Many of these choices pertain to the visual aspects of a text, and therefore had an important role to play in the comprehension and retention of a text, as modern multimodal studies and cognitive psychology have shown (see e.g. van Leeuwen 2004; Levin 1981). In some cases, it can even be argued that visual cues guided the way a text was read out loud, a practice prevalent at all levels of society in the early modern period.

This paper will highlight the functions of typography and illustrations in early English printed texts, with some hypotheses on patterns and changes over time in these features. Examples will be drawn from the *Corpus of Early Modern English Witchcraft Pamphlets* (Suhr 2009) and the *Corpus of Early Modern English Medical Texts* (Taavitsainen *et al.* 2010). The choice of the type used for a text is in itself an indication of its intended audience well into the seventeenth century, for the shift from black-letter to roman type occurred much earlier in “high” literature such as vernacular scientific treatises than in texts aimed at the lower classes, such as ballads and pamphlets (McKerrow 1928: 297). Typographical switches within the texts have a variety of pragmatic functions, from code-switching, quoting, and highlighting important words to textual organizing through headings and marginalia. Illustrations also have various functions, depending on how closely they are embedded into the text (Levin 1981). Understanding the importance of visual features in their contexts and incorporating them into pragmatic analyses of historical texts can thus enhance our understanding of some patterns and changes in English written texts of the early modern period.

Workshops

Workshop: Mergers in English: Perspectives from phonology, sociolinguistics and psycholinguistics

Lynn Clark (Lancaster University)

Warren Maguire (University of Edinburgh)

Kevin Watson (Lancaster University)

It has been claimed that “if linguistics can be said to be any one thing it is the study of categories” (Labov 1973: 342, see also Macaulay 1978, Campbell-Kibler 2009). It is therefore no surprise that phonological mergers, the ‘coming together’ of two linguistic units into a single category (Crystal 2003: 289), has long been a core part of English linguistics. Exploring mergers in English historical phonology has usefully highlighted processes of structural development in the various stages of the language (Maguire, 2007). In studies of English sociolinguistics, mergers-in-progress allow for the investigation of how these sound changes develop, and how they might expand through the speech community and take on social meaning (Labov 2001, Baranowski 2007, Hall-Lew 2009). And in psycholinguistics, mergers offer the potential for questions to be asked about the relationship between linguistic production and perception (see Warren & Hay 2006 and Hay, Warren & Drager 2006 for work on mergers in New Zealand English). While mergers have been well studied from these different individual perspectives, the cross-fertilisation of ideas from one subfield to another is often limited. One reason for this is that discussions typically take place at different venues, each with a different focus. Historical linguists, for example, might discuss mergers at the *International Conference for English Historical Linguistics* (ICEHL), whereas work on current mergers in varieties of English might appear at the *International Conference for the Linguistics of Contemporary English* (ICLCE), and discussions of the sociolinguistics of English mergers might appear at *New Ways of Analysing Variation* (NWAV). This is unfortunate, not only because it adds weight to the claim that linguistics has suffered from a ‘methodological monoism’ (Arppe et al 2010), with different subfields traditionally relying on one type of evidence, but also because in order to fully understand how phonological mergers work, how they spread or recede through a language, or even, fundamentally, what it means to refer to ‘a merger’ in the first place, a range of evidence is required. In this workshop, we propose to bring together perspectives from historical linguistics, variationist sociolinguistics, phonetics & phonology, and psycholinguistics, under the ISLE umbrella, in order to engage with the following 4 questions:

1. What is a merger, and how do we know?

If a phonological merger is the coming together of two linguistic units, then a key question must be asked about how we determine whether two units are the same or different in the first place (Maguire 2007). What evidence do we need to determine whether a particular merger has taken place? A range of different evidence can be found in the literature, such as native-speaker intuitions (e.g. minimal pair test, Labov 1994: 353) judgements gathered from perceptual tests using either authentic or acoustically manipulated stimuli (e.g. the commutation test, Labov 1994: 356), or auditory and

acoustic analysis of the phonetic similarity of the sounds in question (see among much else Watt 1998, Maguire 2007, Baranowski 2007). Whether two sounds are the same or different is even more pertinent given the existence of near-mergers (Labov 1994: 349-370), in which speakers report two sounds as being the same but maintain a consistent phonetic difference between them in production. This consistent difference will be small, and this highlights a related methodological question: what is the best way to measure a merger? Traditionally, at least for vocalic mergers, the distance between the formants of two vowels in question are measured in Euclidean terms and compared. More recently, other measures have been used which take a wider range of variability into account (e.g. in vowel production, the Pillai-Bartlett statistic, see Hay, Warren & Drager 2006; Hall-Lew 2009, and in vowel perception, the Weber fraction, see Rosner & Pickering 1994). Which of these methods offer the best ways of analysing merger? And how can they be used in the discussion of what a merger is?

2. How do mergers develop?

The term ‘merger’ refers to what is essentially a historical process, but in order to explore how mergers develop, including how they spread across speech communities, we need to examine mergers-in-progress. Labov (1994) claims that mergers can develop in three main ways – by transfer, by drift, or by expansion. Understanding which of these mechanisms is at work in the development of a particular merger in a given speech community is important as it allows a number of predictions to be made which in turn affect our understanding of the resulting output. For example, while a merger by transfer is lexically gradual and phonetically abrupt, a merger by drift is phonetically gradient but lexically exceptionless. This means that in a merger by drift – but not a merger by transfer – we would expect to find phonetically intermediate forms between the two sounds in question. How can current methodological practices in sociophonetics help us to determine the type of merger taking place? Another question relates not to the expansion of mergers, but to their reversal. Labov’s (1994: 311) seminal claim is that mergers are irreversible by linguistic means (“once a merger, always a merger”). Later work claimed that mergers are reversible given the appropriate social conditions (Watt 1998, Maguire 2007). So, if mergers can be reversed, what are the appropriate social conditions? And who leads the change?

3. How are mergers evaluated by speakers?

Once phonological changes are innovated and begin to diffuse across speech communities, they are usually said to be in some way sociolinguistically salient (Trudgill 1986). However, there is disagreement about the extent to which mergers in English are subject to social evaluation. Labov (2001: 27) has claimed that mergers in American English are ‘invisible’ to such observation, and so argues that their spread cannot be due to factors of prestige, whilst Warren & Hay (2006) claim that a particular merger-in-progress in New Zealand English is directly commented upon both by speakers in interviews and more widely in the media. Far from being invisible to social evaluation, then, it seems that some mergers are salient enough to be classified as ‘stereotypes’, in the Labovian sense. Whichever position is correct, questions remain about exactly what listeners attend to: do people react to the changing phonetic qualities of the vowels in question or to the collapsing of two previously distinct categories? Or both? And in what

ways? And, crucially, how do we best tap into speakers' metalinguistic knowledge in order to answer these questions?

4. What can mergers tell us about phonology?

With any merger-in-progress, it follows that speaker-listeners will be surrounded by merged and non-merged forms. The status of (socio)linguistic variation in phonology has received minimal attention in generative frameworks but recent work on usage-based and exemplar models of phonology (e.g. Docherty & Foulkes 2000, Foulkes & Docherty 2006, Foulkes 2010, Clark 2009, Warren & Hay 2006, Hay, Warren & Drager 2006) are now attempting to answer question like: to what extent is knowledge of phonological variation part of phonological knowledge? How do phonological and sociolinguistic constraints interact in shaping phonological change? Which phonological models best represent patterns of change? What does this mean for our understanding of phonological systems? The availability of large spoken corpora for some varieties of English (e.g. ONZE, Hay & Fromont 2008) means that work on mergers-in-progress can provide the perfect testing ground for research which attempts to seriously address the synthesis between variation studies and phonology & phonetics. Exploring these overarching questions in the workshop therefore not only allows us to contribute to the discussion of phonological mergers in English, but "goes right to the centre of what we are doing as phoneticians, (historical) phonologists, and linguists" (Maguire 2007: 8).

Participating Abstracts

The meaning of 'merger'

Warren Maguire (University of Edinburgh), Lynn Clark (Lancaster University), and Kevin Watson (Lancaster University)

This talk will frame the key questions to be explored throughout the workshop. First we will discuss the meaning of the term 'merger', which has been applied to a range of different phenomena. We will show that precisely how a merger is conceived has important consequences for our understanding of phonological variation and change. For example, it may be applied to a synchronic state, or to different kinds of diachronic change (merger by transfer, merger by drift, merger by expansion). It may refer to the phonology of an individual, or to a speech community (which may or may not be homogenous with respect to the feature). Furthermore, some mergers are variable even within the phonology of the individual, leading to variation in production or to a mismatch between production and perception. Other mergers are lexically restricted, or involve a high degree of phonetic overlap but not complete identity ('near mergers'). We will show that an examination of these various factors is necessary for a full understanding of 'mergers' and their (ir)reversibility.

On the role of social factors in vocalic mergers

Maciej Baranowski (University of Manchester)

This paper discusses the role of gender, socio-economic status, and style, in the loss of phonemic distinctions. It presents evidence bearing on two of the most important

generalisations emerging from studies of sound change in progress conducted in the last few decades: the curvilinear hypothesis and the tendency of females to lead linguistic change (Labov 2001). Support for those generalisations has so far come largely from studies of vowel shifts and lenition processes; there has been little evidence from large-scale studies of mergers.

This study looks at the *cot-caught* and the *pin-pen* mergers currently in progress in Charleston, South Carolina. It is based on sociolinguistic interviews with 100 informants, aged 8-90, covering the entire socio-economic spectrum of the community. The speech of 90 of the informants has been analysed acoustically. Spontaneous speech is supplemented with minimal pairs and word list reading. F1/F2 measurements and minimal-pair test results are subjected to a series of multiple linear regression analyses, with social class, gender, age, and style as independent variables.

While Charlestonians over the age of 50 distinguish between the two low-back vowels in perception and production, speakers around the age of 20 and younger are usually merged. Speakers aged 25-50 are variably merged. Women show a lead over men and there is a curvilinear pattern for social class. There is an apparent lack of social affect associated with this merger: there is little style-shifting and there are no overt comments from the informants.

In the *pin-pen* merger, on the other hand, multivariate analysis reveals a strong social class effect in addition to age, but no gender difference. The degree of merger appears to be inversely correlated with social class, showing a monotonic, rather than curvilinear, relationship, with higher classes displaying a lower degree of merger. There are also overt comments from the informants, indicating that there is some awareness of the *pin-pen* merger and a possible association of this feature with the inland South. This suggests that while mergers are usually devoid of social affect, some mergers, possibly those associated with a particular region or social group, may be above the level of conscious awareness.

Capturing listeners' real-time reactions to the NURSE~SQUARE merger

Lynn Clark and Kevin Watson (Lancaster University)

In research on language attitudes, the dynamic nature of response elicitation has received little attention (although see Labov 2008). Evaluative responses are dynamic events, constantly changing through time as the signal unfolds, yet the tools we typically use to capture these reactions allow us only to see a snap shot of this process by recording reactions at a single point in time. Continuous audience response data collection tools have been successfully applied to studies collecting dynamic real-time reactions to performing art (Stevens et al 2009), music (Schubert 2001) and dance (Stevens et al 2009). This paper outlines a methodology which attempts to employ similar techniques to capture listeners' real-time reactions to a phonological merger.

While phonological mergers are relatively well studied, there is conflicting evidence over the extent to which they are sociolinguistically salient in a given speech community (cf. Labov 2001 and Warren & Hay 2006). It is unclear whether listeners react to the changing phonetic qualities of the vowels in question or to the collapsing of two previously distinct categories (or both). This paper reports the results of an experimental investigation into listeners' evaluative reactions towards the

NURSE~SQUARE merger in the north-west of England and attempts to shed light on some of these issues.

The north-west of England, namely Liverpool and Lancashire, is a particularly useful location in which to explore these issues because while many speakers have a NURSE~SQUARE merger, its realisation differs across the region: Liverpool speakers typically merge to a front [ɛ:] while speakers from Lancashire merge to central [ɜ:]. To test listeners' responses to each variant, we presented two groups of listeners from each locality with read sentence data from a single speaker. The speaker was from the north-west of England and had a centralised NURSE~SQUARE vowel in his native accent (so was representative of the Lancashire model). To achieve a matched-guise, the original NURSE~SQUARE vowels were acoustically manipulated in order to raise their F2 value to give the impression of fronting (representing the Liverpool model). Listeners from Liverpool and Lancashire were asked to react to whether they thought each guise sounded 'posh', and their reaction was measured in real-time using bespoke audience response software (cf Watson & Clark in prep), administered via the web.

The novelty in this approach is that it can be used not only to examine the evaluative reactions to these guises but also the points in time at which major shifts in response have taken place. This can be done empirically by applying the statistical techniques of Change Point Analysis (Killick 2010). A preliminary analysis of change points in these data suggests that there are significant shifts in evaluative reaction at particular points in time which correlate with instances of NURSE and SQUARE. We present further details of these data in the paper and argue that this is evidence of the salience of this particular vowel merger in England's north-west.

Mergers in production and perception

Katie Drager (University of Hawai'i at Mānoa) and Jennifer Hay (University of Canterbury)

Our primary interest in mergers lies in their ability to shed light on how sounds, lexical information, and social information are stored in the mind and accessed during the production and perception of speech. In this talk, we step through results from several production and perception experiments that focus on mergers. In these experiments, we investigate questions such as: can expectations about a speaker's likelihood to produce merged variants influence identification of distinct tokens produced by that speaker? And are similar tendencies observed across different types of tasks and in speech production and perception?

We have found that New Zealand listeners who produce merged tokens of NEAR and SQUARE can accurately distinguish between the vowels in perception even though they report that they are guessing. The ability to distinguish the vowels is affected by a variety of factors for these listeners, including the likelihood that the speaker and experimenter maintain the distinction. Additionally, we have found that New Zealanders who merge pre-lateral DRESS and TRAP do so to varying degrees depending on whether the word is a real word or a nonsense word and that a token's real word status affects the merger differently in production and perception. These phenomena are currently being explored further with other mergers and in other dialect areas.

The results from these experiments are discussed in terms of how the sounds may be represented in the mind, indexed to non-linguistic information, and later accessed during speech production and perception.

Interpreting 'flip-flop' patterns in vowel mergers-in-progress

Lauren Hall-Lew (University of Edinburgh)

Given a stratified sample of speakers in a particular community undergoing merger-in-progress, occasionally one or two of these speakers will exhibit an unusual phonological pattern known as flip-flop (Labov 1994:143; Hall-Lew 2009; see also Di Paolo 1992:271). In cases of flip-flop, vowels appear to move past the point of coalescence; phonetic difference is maintained, but with the opposite phonetic cues that the vowel classes had prior to the community-wide onset of merger. In this talk I focus on two San Francisco English speakers who exhibit a flip-flop pattern for the LOT and THOUGHT vowels which are undergoing merger in the wider community. Based on an analysis of these speakers' construction of identity in their sociolinguistic interviews, I argue that flip-flop is a potential case of hypercorrection. Community-wide patterns from a sample of 30 speakers suggest that the height and anteriority of the THOUGHT vowel carries social meaning in the community, with higher and backer forms indexing an older authenticity, and lower and fronter productions indexing a new authenticity. In this analysis, flip-flop is one consequence of speakers' self-styling toward new, emergent linguistic practices of local identity.

New contrast acquisition: Methodological issues & theoretical implications

Jennifer Nycz (University of York)

The opposite of merger is phonemic split: when one category becomes two, either in a language variety or in the system of an individual. Splits are less commonly studied than mergers, but raise many of the same phonological issues: for example, the extent to which contrast is a categorical or gradient notion, how contrast is mentally represented, and how changes to the phonemic inventory progress within a speaker or a community.

This paper presents data on the acquisition of a new contrast by native speakers of Canadian English who have moved as adults to the New York City region. I focus on these speakers' realization of the vowels in the PALM, LOT, CLOTH, and THOUGHT word classes - that is, the word classes implicated in the low back vowel merger, "the largest single phonological change occurring in American English" (Labov 1994:316). According to Boberg (2008), "virtually all native speakers of Canada today" have this merger, which has been present in Canadian English for several generations. In contrast, the New York City region is one of a few areas in North America in which the low back vowel distinction remains robust (Labov et al. 2006). The analysis examines how these speakers who natively possess a single low back vowel category have acquired the low back vowel distinction of the new ambient dialect. One finding is that the speakers show remarkable first dialect stability with respect to their low back vowel system, even after many years of new dialect exposure: in minimal pair contexts, nearly all the speakers continue to produce and perceive a single vowel category. However, in word list and conversational contexts, the majority of speakers exhibit a small but significant phonetic

difference between PALM/LOT and CLOTH/THOUGHT, reflecting the separation of these word classes in the new dialect to which they are exposed; moreover, the realization of these words show frequency effects consistent with a lexically gradual divergence of the two vowels.

These findings will be discussed in terms of their implications for theories of phonological representation and change, as well their methodological implications for the study of mergers- and splits-in-progress.

Modelling (socio)linguistic mergers: the role of global context in the processing of social and linguistic information

Phillip Tipton (University of Chester)

Phonological mergers are just one of many potential sources of ambiguity in the speech signal which listeners must resolve in order to accomplish successful spoken word recognition. In the case of the NURSE-SQUARE merger (Wells 1982) in the north west of England, those who have the merger display a complete lack of vocalic contrast in words taken from these lexical sets, though the actual phonetic realisation of this collapsed distinction can vary. This paper considers this merger within the context of psycholinguistic models of speech perception and production and suggests that the general notion of global context effects may prove useful in formally modelling the interaction of social and linguistic information in speech processing.

Much of the work which considers the processing of variation in the speech signal considers either: i) instances of variation which is somehow phonetically unlicensed in the language (e.g. Gaskell & Marslen Wilson 1996, 1998); or ii) variation which is predictable and sociolinguistically unmarked (e.g. Ranbom & Connine 2007). In contrast to these psycholinguistic treatments of variation in the signal have been those studies which report on the dynamic relationship between language and social meaning through an examination of the perception (e.g. Campbell-Kibler 2009; Hay et al. 2009; Drager forthcoming) and production (Drager et al. forthcoming) of (socio)linguistic variables and changes in progress. Implicit, at least, in studies of the latter is the assumption that social information/evaluation plays a role in speech perception, and exemplar models of phonological representation are increasingly cited as a fruitful avenue of research in the goal of accounting for the seeming simultaneity of the processing of social and linguistic information in the speech signal.

In the case of putatively homophonous minimal pairs, cohort models of speech perception (Marslen-Wilson & Tyler 1980; Gaskell and Marslen-Wilson 1997)) predict that the listeners' task is heightened by the need for further ambiguity resolution once the lexical item's uniqueness point has been reached. Although the Distributed Cohort Model (Gaskell and Marslen-Wilson 1997) proposes the parallel processing of phoneme perception and word identification in order to account for the apparent effect of context in Zwitserlood's work (1989) on spoken word recognition in Dutch, in the case of minimal pairs it would seem that context alone must take on the task of ambiguity resolution (though cf. tolerance-based accounts of word recognition which allow all features of a word to be represented in the lexicon e.g. Connine et al 1997). Moreover, if we assume that the lexicon is structured in the way assumed by generative phonologists and many psycholinguists, then a way must be found of accounting for the successful recognition of

non-minimal pair merged tokens by listeners who do not have the merger in production, for whom, one must assume, such a token shares many characteristics of a non-word. In such cases, the potential exists for recourse to both a local (McClelland et al. 2006) and global (Mirman et al. 2008) context account. Mirman et al. (2008) found that global, non-linguistic contexts influence language processing by shifting attention away from contextually-inappropriate meaning. This paper proposes a framework whereby global context acts as both the conceptual and processing bridge between linguistic and non-linguistic perception. In this connection, we argue for a pragmatic-situational account of contextual inappropriateness in which knowledge of social-indexical properties facilitates speech perception.

Workshop: English in the Indian Diaspora
Marianne Hundt (University of Zürich)

English in the Indian Diaspora is a fairly recent area of study. Research has therefore been restricted to individual situations rather than comparisons of language contact across different diasporic settings. The aim of this workshop is to bring together scholars who have been working on the use of English in the Indian Diaspora, with a view to opening up such a comparative perspective and exploring theoretical and methodological questions that may arise. These include:

- To what extent have the different diasporic situations resulted in different outcomes?
- Do we observe a similar set of contact features and processes across diasporic situations?
- Do types of responses to dialect contact in migration help clarify the relative degree of endonormative stabilization of Asian Englishes?
- Is there a relation between the rate of change/adaptation/focusing and the degree of transnational network maintenance or other factors specific to migration?

A thematic focus of this ISLE2 conference is methodology. So an obvious further question is whether the study of language contact in diasporic situations – here, the Indian Diaspora – requires methodological innovations that take us beyond ‘standard’ sociolinguistic methodologies. One area of study where this may be particularly relevant is the mapping of structural features onto the sometimes rather complex set of available social affiliations. What types of ‘ethnic’, ‘local’, and/or multifunctional social indexicalities are achieved by individual speech styles?

The regional spread of Indian Englishes ranges from Singapore via Africa and the UK to the Caribbean and the South Pacific. We hope to include papers both on primary and secondary diaspora situations, covering a wide range of linguistic features and processes of change.

The papers presented at the workshop will address a number of topics/themes. In terms of language structure, they will cover a broad range of features found in the English used in Indian diaspora contexts, namely phonetics (both segmental and supra-segmental phenomena), syntax and lexis. The question we will discuss is whether some features of English in the diaspora are more durable/persistent than others, i.e. will accent features prevail, for instance, or some syntactic patterns be observable even in second generation speakers? Or will most speakers adapt to the new environment over time? Which social factors will foster retention, and which adaptation of forms?

In addition, the papers address the related, socially oriented themes of ideology, identity construction, and types of social embedding: To what extent are diasporic varieties exo- or endo-normative? How do community boundaries and emerging identities affect attitudes, which in turn may influence linguistic form? How internally heterogeneous are diasporic groups (and what are the consequences of these differences?)? And finally, what is the broader positioning of English in the given diasporic setting, e.g. majority language, ESL, part of a multilingual repertoire, or in a shift situation with an English-based creole?

In sum, the workshop offer an innovative comparative dimension to the question of language variation and change — as well as identity formation and social change — in situations of transnational migration.

Participating Abstracts

Indo-Trinidadian speech: features and stereotypes

Dagmar Deuber (University of Münster) and Glenda Leung (University of Freiburg)

As a result of contract labour migration in the 19th century, about 40% of Trinidad's population is of Indian descent; Afro-Trinidadians make up another almost 40%, while the remainder are of mixed or other ethnicity. Trinidadian society today is said to have achieved "a high degree of racial and class integration" (Youssef 2002: 189) and the factor of ethnicity has so far not figured prominently in sociolinguistic studies of Trinidadian English. However, in a language attitude survey by Mühleisen (2001), a majority of respondents claimed that there are ethnic differences in Trinidadian speech.

The aim of the present paper is to investigate the actual and perceived distinctiveness of Indo-Trinidadian speech today, and, on the basis of the findings, to assess the importance of the factor of ethnicity in Trinidad's current sociolinguistic scene. To what extent do differences still exist after the now complete language shift of the Indo-Trinidadian population to English/English-based Creole, and to what extent are they perpetuated as stereotypes?

The paper reports the results of an experiment in which informants were asked to identify speakers' ethnicity on the basis of speech recordings. In this experiment, a majority of speakers were identified correctly as either Indo- or Afro-Trinidadian, but a number were misidentified as belonging to the respective other ethnic group by the majority of respondents. We will try to determine the features and stereotypes that form the basis of these (mis-)judgements by analysing various aspects of the pronunciation of the speakers whose speech was used in the experiment. The analysis will focus on the suprasegmental level (pitch, tone of voice, intonation), which seems to be the main factor, but we will also consider possible differences at the segmental level.

Zero articles in Indian Englishes: a comparison of primary and secondary diaspora situations

Marianne Hundt (University of Zürich)

The omission of articles where British or American English require either a definite or indefinite article is a typical feature of Indian English (IndE). Sharma (2005b) found that substrate influence played a role for the use of indefinite *one* in IndE, whereas pragmatic functions (givenness and modification) played a role in the use of zero articles (see also Sedlatschek, 2009: 227).

Definite *the* and indefinite *a* are also omitted by Fiji Indians in Fiji (Mugler and Tent 2008, Hundt and Schneider, in preparation) and those who have moved to a secondary diaspora in places like New Zealand or Australia. The following example illustrates variable use of articles and zero articles by a Fiji Indian in the secondary diaspora in Wellington, New Zealand:

- (1) coup is Ø only big problem; just because of the coup people left otherwise Fiji was Ø very nice place; otherwise Fiji was a very nice place (first-generation male, in his 70s)

Sharma (2005a) investigated the use of articles by first-generation immigrants from India in the US. Her study shows that zero articles are a feature that is retained in the diaspora

context, even by speakers who are otherwise close to using standard, native-like English. The data for this paper come from fieldwork in Fiji (spontaneous conversations) and sociolinguistic interviews conducted in the Fiji Indian Diaspora in Wellington, New Zealand. Data from the secondary include both first- and second-generation migrants, so it will be possible to investigate whether zero articles are retained in the speech even of people who acquired their English in a predominantly English-speaking environment. It will also be possible to link the use of zero articles to informants' construction of identity in the secondary diaspora. The study will thus combine both quantitative and qualitative methods.

Singapore's Indian community: linguistic, social, and sociolinguistic aspects

Jakob R. E. Leimgruber (University of Freiburg)

The Indian population of Singapore is the smallest of three government-defined ethnic groups, representing just over 9% of the city-state's residents. Compared to the majority Chinese (74 %) and the 'indigenous' Malays (13 %), this category of Singaporeans is highly heterogeneous both linguistically and in terms of actual ethnic and cultural background. While a majority of 54% is of Tamil extraction, as reflected in the selection of Tamil as a co-official language of the republic, the next largest group of Indians are the Malayalis at just 7.6 %. This diversity has its roots in complex migratory patterns often associated with pre-existing class and race distinctions, which, while much less prevalent today, continue to remain visible.

The highly heterogeneous nature of the Indian speech community in Singapore, which also has the highest level of English usage as a home language of the three main ethnic groups, makes for a complex pattern of language use, with several varieties playing a role (English, Singlish, Tamil, other Indian languages/varieties of varying status, Malay). This paper attempts an overview of the Indian population of Singapore, its history, its current linguistic composition, and the language policies that have affected them. Recent research on ethnic varieties of Singapore English (Lim 2000, Suzanna and Brown 2000, Deterding 2007) and ethnic identity marking (PuruShotam 1997, Lai 2004) shows that regardless of the low salience of linguistic features indexing ethnicity, there are important linguistic ways in which 'Indian' identities are marked, not least by the repertoire of codes used in the community. It is argued that the paucity of support towards quantitative evidence for ethnic varieties ought to be redressed by a qualitative reading of language use in the speech community.

The making of a dialect dictionary 1: where does a New English dictionary stop?

Rajend Mesthrie (Cape Town University)

This paper examines the theoretical and practical underpinnings of a *Dictionary of South African Indian English* or SAIE for short (Mesthrie: 2010), a work of about 1600 items characterizing this New English (or social dialect of South African English - SAE). Ultimately, though, the dictionary might be termed a glossary, since it only deals with terms specific to the dialect. This dictionary obviously follows mainstream lexicographical practice (technicalities of format for entries, pronunciation guide as needed, etymologies for all entries from about 10 different languages). It has a

sociolinguist's emphasis on charting salient variation (for example in spellings, as well as usage characteristic of all classes, not just the literate). As such the work aims to document the social history of the dialect and its speakers.

The overriding question facing the lexicographer of such a work is 'where does dialect description stop?' (in relation to the local standard, international standard(s), other related dialects, global culture etc). The paper will discuss decisions taken over the following issues (i) Lexis extremely common in SAIE that is shared with general South African English, e.g. *babelaas* ('hangover'); (ii) Lexis from SAIE that has been adopted more widely in SAE, e.g. *biryani* ('a rice dish cooked with meat and/or vegetables'); (iii) Lexis from Indian languages that continue to be used either widely or in some homes according to ancestral languages, long after those languages ceased to be spoken in the home, e.g. *kaaro* ('chili hot'); (iv) Lexis shared with IE (English of India) and with international English via the British Empire or the new globalization, e.g. *curry*, *Bollywood*; religious terminology from Hinduism and Islam that is shared with India, the Middle East and these religions in diaspora, e.g. *Kavady*, *Ramadan*. There are other complex decisions to be made in relation to geography and slang. The core features of the dictionary come from KwaZulu-Natal province, though decisions have to be made in respect of Afrikaans-English bilingualism in the other provinces. In respect of slang, whilst many speakers believe that forms like *vai* ('to go') and *pozi* ('home') are essential to their dialect, these are in fact part of a largely mainly male form of slang shared with the variety known as Tsotsitaal nationwide. The paper will show that some boundaries have to be set, albeit fuzzy ones, and will explore implications of the overall approach to the Indian diaspora. It will also suggest that in this process dialect dictionaries have their own interests, rather than being purely prescriptive artefacts.

East African Indians in Leicester, UK: phonological variation across generations

Claudia Rathore (University of Zürich)

Recent years have seen an increase of sociolinguistic interest in the use of English in the different strands of the Indian diaspora. Research has been conducted, for example, on South African Indian English (Mesthrie 1992), Indo-Fijian English (Zipp 2010), as well as the variety of English spoken by Fiji Indians in New Zealand (Hundt *fc.*). In the UK, several Indian or, more generally speaking, Asian communities have also attracted scholarly attention, the focus usually being on locally-born generations (see e.g. Rampton 1995, Fox 2007, Khan 2007, Stuart-Smith *et al.* 2009, but cf. Sharma & Sankaran 2009). This paper aims at contributing to this growing body of research by investigating patterns of linguistic variation and change among East African Indians in Leicester, a community of Indian 'twice migrants' who settled in Britain via East Africa in the late 1960s and early 1970s. Like Fiji Indians in New Zealand, East African Indians in Britain present an interesting case study for the linguistic outcomes of dialect contact in double or secondary diaspora situations.

Drawing on data from sociolinguistic interviews, the study focuses on phonological variation across two generations of migrants, with a twofold goal: on the one hand, of exploring how the community's complex migration pattern may have influenced the speech of first generation migrants; and, on the other hand, of determining whether and to what extent linguistic patterns found among first-generation migrants are

transmitted to the locally-born generation. To this end, I undertake a quantitative investigation of two variables, rhoticity (the presence or absence of postvocalic /r/ in *car* or *farm*) and the STRUT vowel (cf. Wells 1982). Variation in the use of postvocalic /r/ is analysed auditorily, while realisations of STRUT are measured acoustically in PRAAT (Boersma & Weenink 2010). The two features are examined at an inter-speaker level and discussed with reference to the varieties of English relevant to the community's sociolinguistic history.

Identity, ethnicity and fine phonetic detail: an acoustic phonetic analysis of syllable-initial /t/ in Glaswegian girls of Pakistani heritage

Farhana Alam and Jane Stuart-Smith (University of Glasgow)

English in the Indian diaspora in Britain is a subject of recent interest, both in the emergence of regional ethnic Englishes (Heselwood & McChrystal 2000) and in the potential links with language change in mainstream English (Kerswill et al. 2008). Also intriguing are the processes by which features percolate through generations of speakers (Sharma & Sankaran 2009).

At the phonological level, features of phonetic interference in the first generation appear with subtly different phonetic characteristics indexing specific identities in second and further generations (Stuart-Smith et al. forthcoming). So in Glasgow, first-generation speakers of Pakistani heritage substitute retroflex plosives for /t d/, but second-generation speakers produce rather different, auditorily 'retracted' stops, which vary according to locally-situated social/ethnic identities (Lambert et al. 2008).

Identifying this kind of phonological development poses some methodological challenges. Auditory analysis yields fairly gross, discrete categories which make strong descriptive assumptions about articulation, e.g. retroflex [ɖ], retracted alveolar [ɖ̠]. But as Lawson et al (2010) show, the auditory-articulatory relationship is complex: Scottish 'retroflex' /r/ often does not show a tongue raising gesture. Again, even trained phoneticians may not be sensitive to fine-grained phonetic characteristics which are nevertheless salient for a local community (Foulkes & Docherty 1999).

Here we present the results of an acoustic phonetic analysis of syllable-initial /t/ in adolescent girls of Pakistani heritage in Glasgow. Speech data were drawn from a long-term ethnography in an inner-city secondary school. Several Communities of Practice were identified (see Eckert 2000), spanning a continuum from British/western to traditionally Pakistani and/or Muslim practices. A spectral analysis of the stop bursts of /t/ reveal clear patterning according to social practices, but also gradience for individuals within Communities of Practice. This suggests the emergence of a local ethnic accent with the subtle adaptation of heritage features reflecting new emerging identities (cf Harris 2006).

Features of Indo-Fijian English across registers

Lena Zipp (University of Zurich)

Among the various countries of the Indian diaspora, Fiji presents a complex sociolinguistic scenario that is in some points similar to the South African situation described by Mesthrie (1992). Going back more than 130 years to the times of indentured

labour, the history of Indians in Fiji is shaped by socioeconomic achievements as well as by ethnical tensions with the indigenous Fijian population, and continuing political unrest leaves the Indo-Fijians of today with mixed feelings towards their Pacific home country. It is to be expected that political developments like the 2000 and 2006 coups and the 2009 constitutional crisis in the recent history of Fiji have their repercussions in the population dynamics and thus also influence the sociolinguistic language spectrum encountered among Indo-Fijians; particularly at the upper end of the socioeconomic scale, where greater mobility, increasing emigration and high transnational network maintenance potentially lead to a re-evaluation of linguistic identities. This development could take two possible directions: either towards a shared, national, endonormative model of second language English, i.e. enforcing local norms, or towards a regional or global exonormative model, such as established first and second language varieties. Furthermore, there is the possibility that both local and regional/global models are operative depending on the register or genre investigated.

The paper proposes to examine the stylistic variation of selected lexico-grammatical and prosodic features in Indo-Fijian English, based on data collected for the *International Corpus of English* (ICE), Fiji component. The data sampled for the ICE project stems from educated speakers and comprises various registers of spoken and written language, including the original sound files for spoken data. Considering formal as well as informal registers of Indo-Fijian English, it will be investigated whether exonormative models outweigh endonormative ones, or exist alongside each other. In addition to this corpus-linguistic method, the study will be complimented by attitudinal questionnaire data concerning stylistic variation in English among Indo-Fijians.

**Workshop: Aspects of Methodology and Pedagogy:
Long-term research projects on local varieties of English**
Lars Hinrichs (University of Texas at Austin)
Stefan Dollinger (University of British Columbia at Vancouver)

Linguistic research on the speech of the surrounding community has been part of the curriculum at many universities, particularly public institutions, for a long time. New digital tools provide ways of increasing efficiency in this effort by broadening our empirical bases and developing new pedagogical routes toward involving students in data collection and analysis at high levels of quality. Being directed at phenomena that are very much of the real world, such research provides ideal grounds for involving students of all levels in professional academic work.

Numerous publications have emerged from such projects, drawing on data and (sometimes) results that are produced by students, and some universities have declared the inclusion of a research element in their undergraduate curricula a prime goal. Students act as conduits into social networks (Hoffman and Walker 2010, Bigham et al. 2010), conduct dialectological fieldwork, and collect material for database projects (Kretzschmar et al. 2006). Software tools are being developed to facilitate a number of these tasks (Kendall 2008, Purnell et al. 2010, Dollinger 2010). These changes in teaching culture are changing research and funding culture, as well as, quite possibly, the role of the professional linguist.

The presentations in this workshop discuss how research projects on local varieties of English can be set up and directed in ways that are effective and sustainable in both the academic and the pedagogical sense.

Participating Abstracts

**The Theoretical and Methodological Challenge of Longitudinal Studies:
The Case of African American English**

Walt Wolfram (North Carolina State University)

For two decades now, an unparalleled longitudinal study of the development of African American English has been undertaken by the Frank Porter Graham Child Development Institute and the North Carolina Language and Life Project. The project started in 1990, following a cohort of 70 working-class African Americans progressively for 17 years. In the process, more than 2,500 CDs of data were collected on language use, family life, educational achievement, and social environmental—from birth through the completion of secondary education. We discuss how data collection, management, and analysis have developed and been maintained in order to ensure sociolinguistic validity and reliability while preserving data from a longitudinal study that will never be duplicated. Issues of archiving, analysis, and presentation are considered, along with some emerging empirical findings of the study that counter unjustified assumptions about the development of AAE during early childhood and adolescence.

The Wisconsin Englishes Project and WiSCO

Thomas Purnell, Eric Raimy and Joseph Salmons (University of Wisconsin, Madison)

The Wisconsin Englishes Project (WEP) has developed a software tool, Wisconsin Speech Chain Online (WiSCO), that supports the teaching of linguistics, data-based research and field methods of data collection. WiSCO allows individuals to upload acoustic data to a data base, to visualize their own data and to compare their data to other data in the database. WEP combines the use of WiSCO in teaching linguistic courses to create a synergy among introducing students to dialectal variation, data collection and basic research. We will demonstrate the main features of WiSCO and discuss how it has played a role in contemporary research on Wisconsin Englishes.

Student Participation in the Linguistic Atlas Project

Bill Kretzschmar (University of Georgia)

Over time students have become much more involved in the Linguistic Atlas Project (LAP). Graduate students first became analysts of LAP data in their dissertations, and then became collectors of the data, too, as Lee Pederson and Roger Shuy at Chicago, and later many students at Emory and Georgia, executed interviews. Most recently, the LAP at Georgia has given research opportunities to a great many undergraduate students involved in keyboarding old paper records, and in processing audio-taped interviews to create digital versions for the archive and for public use. These students see firsthand how collaborative research works, and also get to use the research products in undergraduate classes. In our American English course, e.g., use of interactive mapping and data lists from the LAP Web site is a key pedagogical technique, from which students learn about geographical and social distributions in our survey data.

Goals for the project and your career: Long term success

Kirk Hazen (West Virginia University)

Setting and re-setting realistic goals is both one of the most important and most ignorable components of your career. Yet to maintain a successful dialect project, the goals for your career and for the project must be tightly integrated and frequently reviewed. This panel presentation argues that reconceptualizing your career will provide maximum benefits. Professional linguists might perform numerous jobs, including professor, administrator, research director, expert witness, and linguistic consultant. The essential turn must be to view oneself as a professional linguist apart and above the roles you play in your primary job. The self-realized status as professional linguist provides the biggest conceptual umbrella under which one can tuck all those other duties. The organizational scheme of a dialect project works much the same way in that your duties as a professor (teaching, research, service) should be incorporated with the dialect project's goals. This presentation also discusses balancing public interest in dialects with one's research interests and how to work with limited resources.

Workshop: Aspects of Methodology and Pedagogy: Evaluation and Instruction
Marnie Reed (Boston University)

Participating Abstracts

Variation in the production of the TRAP vowel in advanced Polish learners of English: Beyond averages

Jarosław Weckwerth (Adam Mickiewicz University)

Polish has two vowels, usually symbolised /a/ and /ɛ/, in the area where English has three – TRAP, DRESS and STRUT (Jassem 2003; Wells 1982). As a result, virtually all theories of the acquisition of L2 pronunciation will predict problems for Polish learners of English (and for learners from many other L1 backgrounds lacking the contrast), differing only in the details of their predictions. This paper will illustrate one area that has been largely neglected in L2 literature (albeit see e.g. Jongman and Wade 2007 or Morrison 2003) – the scope and nature of variation in learner productions.

A corpus was collected containing recordings of read and semi-spontaneous English speech from 150 Polish students of English. The participants were first-year students newly enrolled in an “English studies” programme at a Polish university. Thus, recordings were obtained of advanced students before they received substantial systematic training in English pronunciation (which is part of the programme in the first two years). The present paper discusses results from a subset of the corpus: four repetitions of a specially designed wordlist from 50 female students. The recordings were made in a sound-treated room using a condenser microphone connected to a PC computer via a USB audio interface. The sound files were annotated in Praat (Boersma and Weenink 2010); measurements of F1 and F2 taken at the midpoint of each vowel were subsequently normalised according to the Nearey model (Nearey 1977) using the NORM normalisation suite (Thomas and Kendall 2007).

While aggregate group results showed a statistically significant difference between the average F1 and F2 values of TRAP and both DRESS and STRUT, and the distributions of DRESS and STRUT did not overlap, the distribution of TRAP was “superimposed” on those of the other two vowels. Inspection of individual speakers revealed that all logical “assimilation” possibilities were attested. Cluster analysis showed that some speakers displayed a (more or less) separate TRAP category; some merged it with DRESS; others with STRUT; and, most importantly, there was also a group of speakers for whom TRAP showed a “bimodal” distribution, with some instances in the DRESS area, and some in the STRUT area. There was also considerable variability between lexical items (also inter-speaker).

A number of (tentative) conclusions can be drawn: (1) Presenting results from multi-speaker studies as averaged values can be somewhat misleading, especially with learner subjects, as it tends to hide variability; the application of sociophonetic methodologies, as exemplified e.g. by Labov et al. (2006), is welcome. (2) Results of the kind discussed above are not easily accommodated by extant theories, even sophisticated ones, such as Flege’s (1995) Speech Learning Model. (3) One explanation could be that the variability seen in EFL subjects may be due to variability in the input, which is greater than in usual immersion settings on which much of EFL phonetics/phonology research is based, as suggested by Bohn and Bundgaard-Nielsen (2009); however, this is difficult to assess as it is nearly impossible to form an accurate picture of learner history in EFL settings.

A morphosyntactic algorithm for sentence building in language acquisition

Izabela Lazar (University of British Columbia)

“Grammar is the structural foundation of our ability to express ourselves. The more we are aware of how it works, the more we can monitor the meaning and effectiveness of the way we and others use language. It can help foster precision, detect ambiguity, and exploit the richness of expression available in English.” (David Crystal)¹

One of the most inhibiting circumstances experienced by learners of English consists in their limited control over the sentence construction process. This paper proposes an algorithm for building sentences in English, algorithm which may allow writers to operate word sequences accurately when composing a sentence. As its direct focus has been generated by a sense of urgency (within the domain of language acquisition), this paper is intended as a synoptic platform, where elements of formal grammar (with its reliability on rules and patterns) are intertwined with functional tools, or with choices that must be meaningful in context. (Halliday 1994) It also echoes aspects of lexical functional grammar (Bresnan and Kaplan 1995), although its emphasis is mainly on syntax and morphology. In its initial stages, the method I describe here requires supervision and consolidating explications on the part of the instructor; however, once the learners have acquired the algorithm, they will be able to take control, progressively, of their own writing, as they continue to add subsequent layers of complexity to their sentences. The principle underlying the method relies on using (known) sentence patterns at simple sentence level; this paper will bring three of these patterns for illustration:

- Pattern 1 (*Subject+Verb+Direct Object/S+V+DO*)
- Pattern 2 (*Subject+Linking Verb+Complement/S+LV+C*)
- Pattern 3 (*Subject+Verb+Indirect Object+Direct Object/S+V+IO+DO*).

In essence, a sentence pattern is comprised of a fixed sequence of sentence parts, here named ‘units’ (for discrete words and short phrases) or ‘units of thought’ (for longer phrases and clauses). In turn, each ‘unit’ can be expressed through a finite number of options, here called ‘unit options’. For example, Pattern 1 (P1) contains three units: *Subject*, *Verb*, and *Direct Object*. Each unit can, then, be expressed as follows: the *Subject* options can be one-word nouns, compound nouns, pronouns, noun phrases, noun clauses, infinitives, or gerunds. The *Verb* options can be one-word verbs, specific verbal forms for tenses/aspects/modality/voice (or containing auxiliaries), modals, verb phrases, or idiomatic verbal expressions. The *Direct Object* unit options can be (similar to the Subject’s) one-word nouns, compound nouns, pronouns, noun phrases, noun clauses, infinitives, or gerunds. When referring to sentence parts as ‘units of thought’, I rely on elements of word grammar and word typology (Richard Husdon 2010) and will show that stable principles of syntactic organization can allow for a more specific and accurate use of the sentence patterns.

When introduced to the students, this algorithm will be used in two stages: first, by inviting students to identify a pattern in a-priori-constructed sentences, and second, by

¹ David Crystal, "In Word and Deed." *TES Teacher*. April 30, 2004.

directing learners to build similar constructions themselves and to place them in a delineated format, such as a table. Each table will contain the syntactic units and will be preceded by the morphologic categories required by each specific pattern. For example, Pattern 1 for simple sentences will be expanded in the following format:

S	V	DO
The girl (noun phrase)	drinks (verb)	tea. (noun)

In addition to the patterns of simple sentences, the notions of compound, complex, and compound-complex sentences will also be considered when modelling a pattern. Nevertheless, it is important to note that under this algorithm, the coordination and subordination processes are contingent on the sentence patterns principle. For instance, in this paper, the coordinating and subordinating combinations for one pattern will first use the very same pattern used for simple sentences. To illustrate, I will list a string of derivations for Pattern 1 (P1):

IC = independent clause

ICP1= independent clause – Pattern 1

ICP2= independent clause – Pattern 2

DC= dependent clause

DCP1= dependent clause – Pattern 1

DCP2= dependent clause – Pattern 2

Simple Sentence:

S + V + DO(P1)

The girl drinks tea.

Compound Sentence:

ICP1 + ICP1

The girl drinks tea, and the man reads a book.

Complex Sentence:

ICP1 + DCP1

The girl drinks tea when the man reads a book.

DCP1 + ICP1

When the man reads a book, the girl drinks tea.

Compound-Complex Sentence:

ICP1 + ICP1 + DCP1

The girl drinks tea, and the man reads his book if he can find it.

Therefore, all these clauses use Pattern 1. Still at Pattern 1 level, one could also use any of the combinations of unit options described on page 1.

As more sentence patterns are then introduced, the compound and complex combinations will reflect this. Thus, after introducing Pattern 2, the compound and complex derivations for Pattern 1 will be revisited and expanded to reflect more patterning possibilities:

Compound Sentence:

ICP1 + ICP2

The girl drinks tea, and she is happy.

Complex Sentences:

ICP1 + DCP2

The girl drinks tea when she is happy.

DCP2 + ICP1

When the girl is happy, she drinks tea.

ICP1 + DCP2 + DCP1

The girl drinks tea when she is happy and if the man reads the book.

ICP1 + DCP1 + DCP2

The girl drinks tea if the man reads the book and when she is happy.

Compound-Complex Sentence:

ICP1 + ICP1 + DCP2

The girl drinks tea, and she washes the cup when she is happy.

ICP1 + ICP1 + DCP2 + DCP1

The girl drinks tea; and she washes the cup when she is happy and if the man reads the book.

Following Bybee, Perkins, and Pagliuca (1994), I also consider relevant revisiting elements of generative grammar, which allow for the syntactic rules to coordinate the sentence patterns and the morphological choices that satisfy these patterns. The encompassing project under which this algorithm is carried out entails a much broader approach, with a virtually-exhaustive variety of sentence patterns and including a detailed taxonomy of the parts of speech (which will be compiled in an accompanying lexicon organized primarily on morphological principles).

During the next ten months, I will conduct a research study on a working corpus of approximately 1,000 sentences collected from current and prospective learners. The dynamic aspect of the corpus is essential for this project: on the one hand, as I assist students through their writing stages, I will try to identify and address limitations that may occur in the application of the algorithm, and on the other hand, I will be able to invite students to correlate particular patterns with the corresponding sentences they had built, in order to immediately verify the degree of effectiveness of this feedback mechanism.

A Case of Lexicalization: from Middle to Modern English

Do so and Verb Phrase ellipsis in the Canterbury Tales Joanna Nykiel (University of Silesia and Stanford University)

The English *do so* construction is well-known to share certain properties with verb phrase ellipsis (VPE). In an early proposal by Hankamer and Sag (1976), both are characterized as surface anaphora, that is, as constructions that require a linguistic antecedent and one of the same syntactic form. Although syntactic identity is too strong a constraint, the fact remains that the *do so* construction and VPE show a preference for linguistic (discourse-old) antecedents, an absolute preference in the former's case (Kehler and Ward 1999).

Additionally, a semantic constraint has been proposed for the *do so* construction: it selects a dynamic predicate as its antecedent (Kehler and Ward 1999, Huddleston and Pullum 2002, Culicover and Jackendoff 2005). Recent evidence shows, however, that stative predicates may appear as antecedents as well (Houser, submitted).

I argue, based on data drawn from the *Canterbury Tales* (*Corpus of Middle English Prose and Verse*), that the *do so* construction and VPE have different characteristics, and different distributions as well. This early *do so* has developed into a construction separate from, though related to, VPE, where the verb *do* and the adverb *so* have lexicalized as a unit in the sense of Brinton and Traugott (2005). At the same time, the data indicate that *do so* can stand in for non-dynamic predicates early on.

In my data, *do so* appears in embedded and main clauses, takes split antecedents, and can access antecedents from up to nine lines of text prior. All antecedents are discourse-old, dynamic and agentive, but elsewhere in the Middle English corpus, stative antecedents are possible (1):

- (1) First sal ye **luue** god wid al yure herte and wid al yure saul and wid al yure uertu, And ti prome als ti-self; **sua** ah ye at **do** (a1425 *Ben.Rule(1)* (Lnsd 378).

The adverb *so* (with the meaning of *in the aforesaid manner* (cf. *Middle English Dictionary*)) can either precede or follow the verb *do*; they can also be separated by other adverbs (e.g. *right*, *namoore*) or the negative particle. In one case, *do* is used transitively and followed by *so*. This suggests that *do* and *so* do not yet form a unit.

To the extent that VPE is present in Middle English, it appears exclusively in comparative and conditional clauses, with antecedents (dynamic and stative alike) found in the immediately preceding clauses. Interestingly, if *do* is used as a proform following modals or the auxiliary *have*, coordinated clauses and other embedded ones are attested. There is evidence that the functions of VPE and *do so* overlap, though. I found one clause where manuscripts differ with respect to the construction used. Consider (2-3), where the antecedents are agentive and located in the coordinated clause.

- (2) VPE: Assaie and he shal fynde it tat he **dot**. / ffor be we neuer so vicious wit-Inne (14c. Chaucer *CT* (Petworth ms)).
(3) *Do so*: Assay / and he shal fynde it / tat **sodoth** / ffor be we neuere / so vicious with-Inne (14c. Chaucer *CT* (Cambridge Gg.4.27 ms)).

What this shows is that the two constructions are related but VPE is more restricted in its application.

Given that the modern *do so* construction has its source in the Old English main verb *don* accompanied by the manner adverbial *swa*, which required agentive antecedents (Higgins 1992), there has been retention of the early properties following from the presence of the adverb *so*, but not of the verb itself. This is suggested by the fact that non-dynamic predicates are attested already in Middle English, though non-linguistic antecedents are not. Further, because the verb and adverb of modern *do so* are inseparable, my results provide evidence for the process of lexicalization.

**Workshop: Global English:
Contact-linguistic, typological, and second-language acquisition perspectives**

Markku Filppula (University of Joensuu)

Juhani Klemola (University of Tampere)

Anna Mauranen (University of Helsinki)

The rise of English to a position of a *lingua franca* in the present-day world has led to the emergence of new varieties that have variously been called ‘New Englishes’, ‘World English’ or ‘International English’. Regardless of the term used, one can discern a group of structural and discoursal phenomena that characterise both these new and ‘old’ varieties to varying degrees. These are phenomena that can best be investigated under the heading of ‘Global English’. They can be found in the use of English in many different contexts by people with differing backgrounds: as a means of communication in international business, politics, science, and education, as well as in the daily communication of hundreds of millions of people in countries where English has obtained the position of a second official or unofficial language. They also occur in the speech of even more people all over the world when they learn English as a ‘foreign’ language as part of their general education.

These developments have not come about without an impact on the English language itself, which is currently undergoing processes of accelerated change, divergence and convergence in several core domains of its grammar and lexis. This comes as no surprise given the unprecedented complexity of the situations in which English is now used and the fact that virtually every major language in the world is in contact with English in some form or other. These processes have in recent years begun to attract the attention of linguists, but so far at least, research has been rather narrow in scope and has focused on either current trends in the mainstream varieties of English, most notably British and American English (e.g. Leech and Smith 2006; Mair 2006), distinctive traits of the ‘New Englishes’ or ‘World English(es)’ which have developed mainly in the former British colonies in East and West Africa, South and East Asia, and the Caribbean (e.g. Görlach 1998, 2002; Trudgill 2004; Bao 2005; Schneider 2007; Mesthrie and Bhatt 2008), or the typologies of standard and non-standard varieties (Kortmann et al., eds., 2004, Filppula, Klemola and Paulasto, eds., 2009). Similarly, although there is a growing body of research into the role and characteristics of English as a *lingua franca* (see, e.g., e.g. Firth 1996; Jenkins 2000, 2007; House 2002 and 2003, Foley 2006; Seidlhofer et al 2006, Mauranen 2005, 2006, Prodromou 2008), much more work needs to be done to find out the exact relationship between *lingua franca* phenomena and the ongoing changes in other varieties of English. As for studies of language change and language contacts, this line of research has to date had little to say about the kinds of change that are taking place in English now (e.g. Traugott 2007), despite the fact that they are likely to have an effect on its basic typological characteristics. The recent developments in the English tense and aspect systems are but one example of such a typological change (see, e.g. Gachelin 1997 and Smith 2005 on the expansion of the English progressive beyond its traditional uses), and other typological realignments can be expected to be under way.

This workshop sets out to examine these recent developments in English through a combination of three methodological and theoretical approaches that we believe are

essential to a better understanding of the processes that English, in its spoken forms in particular, is undergoing as a world language, viz. contact-linguistic, typological, and second-language acquisition (SLA) research. Through combining these three viewpoints the talks in the workshop will examine some of the major ongoing developments in English, their sources and manifestations in different kinds of contexts, as well as their typological implications. A common methodological denominator in the talks will be corpus-based comparison between selected established, or 'standard', L1 varieties and their nonstandard varieties, on the one hand, and these and the new L2 varieties of English, including *lingua franca* English, on the other. The focus of the papers in the workshop will be on major morphosyntactic and discoursal phenomena common to the uses of English in both L1 and L2 contexts and the major contact-linguistic, typological, and SLA-based factors explaining them.

The papers presented in the workshop will seek answers to the following research questions and try to find evidence to test the following hypotheses:

(i) *What features can be discerned in the current developments of English as a global language and how 'global' are they?*

We can expect to find commonalities in morphosyntactic, lexical, and phraseological features in different kinds of present-day English used in a wide variety of settings. On the basis of our preliminary findings and those mentioned in previous research by others, these include the extended use of the *-ing* form even with stative verbs; increased use of the definite article, e.g. with abstract or generic nouns and omission in premodifying descriptive noun phrases; reorganisation of the system of central modal auxiliaries and auxiliation of semi-auxiliaries and modal idioms; use of inversion in indirect questions, and frequent topicalisation with fronting. Similarly, a number of shared functional aspects of use, such as pragmatic and discourse features, can be expected to surface and find forms of manifestation that are independent of particular varieties. For example, discourse markers, metadiscourse and evaluative expressions are used in all kinds of interactional exchange, and can be regarded as vital functions of language. To what extent the distribution of the expressions used for these and patterns of variation in their forms differ across varieties can now be investigated empirically.

(ii) *What mechanisms of actual or potential change can be observed in negotiating meaning and form in face-to face interaction as English is being used by participants from different backgrounds?*

Linguistic group norms can be assumed to arise and be settled in continual negotiations to enable participants to carry out their business without undue hindrance from the language used. Some group norms will be transitory and local, but close observation of the process of their formation helps to understand micro-level mechanisms involved in contact and change; for example, accommodation processes in spoken interaction give rise to convergent linguistic usage, which may have more permanent effects on speaker repertoires when repeated.

(iii) *To what extent can the observed developments be explained by general models and typologies of contact-induced change, including those established for SLA contexts?*

Again, previous research suggests that in conditions of intensive and long-standing contacts between speakers of differing linguistic backgrounds, the types of process mentioned above are expected developments.

Participating Abstracts

Varieties of English and language typology

Peter Siemund (University of Hamburg)

Within a functional-typological approach, we consider varieties of English linguistic systems in their own right, investigating the emerging patterns of variation, and, in particular, possible limits on them. Such typological comparisons across varieties are rewarding, but also pose methodological challenges. As many varieties – especially the so-called New Englishes – emerged through language contact and shift, explanations for the observed phenomena are typically sought in parameters such as typological or genetic proximity, frequency, semantic transparency, complexity, markedness, universal acquisition strategies, but also in system-internal variation of the languages involved. It has also been hypothesized that some grammatical domains are more vulnerable to contact-induced change than others.

In the first part of my presentation I will provide a topical report on current approaches to varieties of English in the functional-typological paradigm trying to categorize the existing research work.

In the second part, I will introduce a few case studies that illustrate important methodological problems for a typological approach and develop ideas how to cope with them. These concern the internal differentiation of varieties (acrolect, mesolect, basilect), distributional restrictions of variety-specific phenomena and their usage contexts as well as so called ‘shared features’ that have been discussed in terms of notions such as angloversals or vernacular universals. I will illustrate these points by drawing on evidence from the areas of interrogative inversion, the present perfect, case-marked pronominal forms, and discourse markers.

Lexical and accent accommodation in ELF interaction

Niina Hynninen and Henrik Hakala (University of Helsinki)

Given the proactive nature of English as a lingua franca (ELF) communication in that its speakers seem to acknowledge the importance of adjusting their language use to enhance mutual understanding, ELF contexts provide rich data for analyzing linguistic accommodation. A focus on such micro-level mechanisms helps us understand the principles of ensuring intelligibility between participants. Further, by exploring the ways participants accommodate to each other rather than conform to some ideal notion of correctness, we can find out about the importance of accommodation in influencing social relations of the participants.

This presentation focuses on accommodation processes in interrelated speech events where participants use ELF. The data come from English-medium higher education events in Finland. We explore both short-term and long-term data concentrating on convergent linguistic usage indicative of group formation. The analysis

looks at the ways in which lexical and accent accommodation reflect the development of social relations between the participants.

Syntactic features in global Englishes: How ‘global’ are they?

Heli Paulasto (University of Eastern Finland), Elina Ranta (University of Tampere) and Lea Meriläinen (University of Eastern Finland)

Recent decades have witnessed an increasing interest in nonstandard or vernacular features of English attested in varieties spoken around the globe, whether in the Inner, Outer or Expanding Circle. In L1 and L2 varieties of English, such features have been ascribed varying, e.g. to endogenous development, contact influence, ‘universals’ of vernacular English or cognitive processing (e.g. Filppula, Klemola & Paulasto 2009a). In English used by EFL (English as a foreign language) learners, similar features have typically been characterized as ‘errors’ resulting from either native language influence or universal processes of second language acquisition (e.g., Swan & Smith 2001). However, research into learner English has not yet sufficiently been brought into contact with research into Inner and Outer circle varieties of English in order to ascertain whether these nonstandard features are similar in L1, L2 and learner varieties of English and, hence, indicative of a development that is taking place in Englishes throughout the world. This paper explores the question of whether these nonstandard features are an indication of the birth of a new Global English, or whether each variety should be considered in the context of its linguistic ecology and individual speakers’ characteristics as English speakers or learners.

The present study focuses on two widespread nonstandard features of syntax, the extended uses of the progressive form and embedded inversion (e.g., Platt *et al.* 1984, Römer 2005, Ranta 2006, Hilbert 2008, Filppula *et al.* 2009b, Sand and Kolbe 2010), in a wide variety of corpora. On one hand, these features are examined with a view of the whole, i.e. the commonalities that can be found in L1, L2, ELF (English as a lingua franca) and learner English data, and on the other, the background of each variety is considered as the potential explanation for the feature. Our data include ICLE, ELFA, MICASE, SED, ICE-GB, ICE-Ireland, ICE-India, and regional Welsh and Irish English corpora. The data analysis shows that these nonstandard features occur widely in different varieties of English, but apparent similarities also conceal complex structural, functional and distributional patterns of variation. Through these corpora, we illuminate these two features of syntax from several angles to see how ‘global’ these features are and what types of variation they exhibit in different forms of English.

Tracking down American impact on Asian and Pacific Englishes in electronic corpora

Edgar Schneider (University of Regensburg)

As is well known, except for the Philippines all of the Postcolonial Englishes of the Asia-Pacific region can be traced back to British colonialism, and in almost all countries an external British norm is still upheld as the target of language education (cf. Schneider 2007). However, an increasing impact of American English has been observed and commented on in many countries, and it is usually met with ambivalent attitudes.

Evidence for this perceived Americanization is scant and anecdotal, however, and serious investigations of the process are missing. This paper sets out to systematically investigate this phenomenon, using corpus data drawn from the International Corpus of English (ICE) project (<http://ice-corpora.net/ice>). Special emphasis will be paid to methodological problems of identifying, detecting and measuring American impact in different contexts.

First, it is necessary to define the notions of Americanization and globalization and, in particular, to develop a method for detecting and weighing American impact. Of course, there is a set of well known, apparently simple distinctions between British and American English on the levels of lexis (e.g. *petrol* – *gas*, *autumn* – *fall*), grammar (e.g. *got* – *gotten*, *Have you ...* – *Do you have ...*) and orthography (e.g. *colour* – *color*), in addition to phonological differences (not considered here, given that this study is based on written data). Recent research (Algeo 2006, Mittmann 2004, Rohdenburg & Schlüter 2009) has shown, however, that reality is much more complex: differences between the two major varieties of English are rarely so clear-cut. Instead, they consist of a vast array of subtle quantitative preferences of either variety for certain lexical choices, phraseological collocations, and grammatical patterns. One question asked, thus, is whether the putative global impact of American English manifests itself more strongly, or possibly even restricts itself, to a few salient features, or to what extent subtle co-occurrence preferences are also influential, if at all.

Based on the above studies by Algeo, Mittmann, and Rohdenburg & Schlüter, a set of word choices, "lexical bundles" (after Biber et al. 1999) and structural preferences distinguishing British and American English are selected, and their frequencies of occurrence in a range of Asian and Pacific Englishes are investigated and interpreted in the light of the putative Americanization process. Specifically, ICE-GB (from Great Britain) and a roughly similarly composed corpus of American English texts (culled from the FROWN and Santa Barbara corpora, given that no ICE-US is available) serve as quantitative benchmarks, and the one-million word ICE corpora from India, Singapore, Hong Kong, the Philippines, and New Zealand, representing the respective varieties of English, are screened for these variant forms. Distributions are investigated quantitatively. Factors considered include variety; language level (lexis, phraseology, syntax, spelling); style level (spoken vs. written); and saliency (salient vs. non-salient variants). Thus, the study yields a subtly differentiated picture of the variant degrees of the impact of American English on Asian and Pacific Englishes.

Question formation in Indian English and in other Southeast Asian varieties

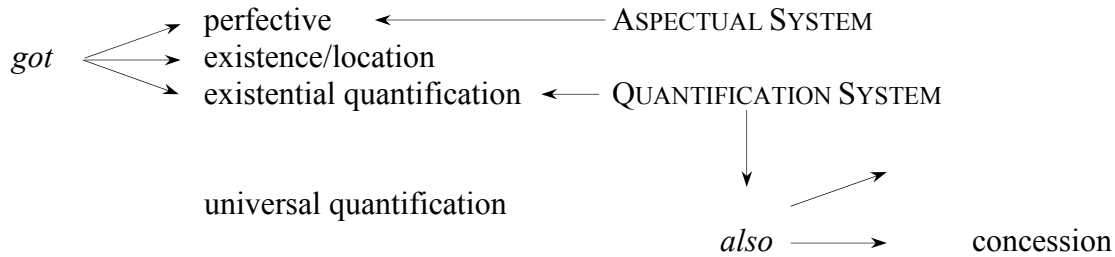
Hanna Parviainen (University of Tampere)

Although previous studies indicate that Indian English has influenced the lexicon of other Southeast Asian Englishes, the actual scope of the variety's influence on the region has not yet been extensively studied. Thus, the aim of this paper is to examine this issue further by studying if the influence of IndE can be detected in the question formation patterns of other Southeast Asian varieties of English. Some characteristics of IndE questions are, for example, the lack of subject-auxiliary inversion in direct *wh*-questions such as "Why you didn't like it initially?" (ICE-IND), the use of inversion in indirect *wh*-questions such as "That's all what can we say" (ICE-IND) and the use of the invariant tag *isn't it* as in "They had come in a bus isn't it?" (ICE-IND). The purpose of

Like its Chinese source *dau*, *also* expresses the concessive function as well:

(5) Even if you pay they *also* don't want to sell

We display the clustering in the following schema:



The web of *got*'s grammatical functions derives from Chinese through systemic transfer.

Diamonds, gender and strong verbs: a study of contact and sociolinguistic factors in the evolution of a variety of Black English in Kimberley, South Africa

Rajend Mesthrie (University of Cape Town)

The present study emanates from a large scale dialectological enquiry into the regional and social differentiation of South African English. It focuses on a small part of the data base, the particular characteristics of working class speakers of Tswana L1 background in Kimberley, a city associated with the discovery of diamonds in the 19th C and with an English culture, plus Afrikaans and Tswana. Black South African English has been reasonably well studied in terms of its phonetics (e.g. Van Rooy 2004) and syntax (e.g. De Klerk 2006). The tendency has been to study the L2 variety of students at colleges and university. Rather less is known about English used by adults in work situations (though there is a small literature from the perspective of ideology and dominance, e.g. De Kadt 1998). While initially interested in regional accent variation, this study focuses on an unexpected feature not reported in the previous literature: viz the treatment of strong verbs by reasonably fluent working-class speakers. Several speakers exhibit a decided tendency towards morphological regularisation, turning strong verbs into weak. Thus words like *digged*, *swimmed*, *feeded* are not uncommon in the data base. The paper presents a quantitative analysis per speaker (no. of speakers = 10) of the use of past tense verbs, showing that there is a gender effect: women use a significantly higher proportion of regularised strong verbs than males in the sample. The study attempts to tease out to what extent contact and psycholinguistic factors are involved (e.g. degree of education in English and exposure to English as target language) and to what extent sociolinguistic factors like social networks and gender. There is an anthropological literature on gender differentiation in language contact affecting Southern African languages, notably the study of the partial and then total shift from Tonga to Zulu amongst a group of Mozambiquan migrants (Herbert 2004). This literature suggests that socio-symbolic functions are served by women's less enthusiastic and less wholesale adoption of the incoming language Zulu than by their male Tonga counterparts. Such an explanation will be weighed against the more mundane one of fewer opportunities of social interaction available to working-class women. In terms of core methodology the paper rests on interviews in the Labovian urban dialectology tradition, adapted to the situation of L2 speakers. The analysis of the weak and strong verbs is quantitative, and the interpretation of the results is socio-symbolic. Ultimately the paper is meant to contribute to the analyses of variation in World Englishes.

Workshop: Current methods in English diachronic linguistics: How can new corpus-based techniques advance historical description and linguistic theory?

Hubert Cuyckens (University of Leuven)

Martin Hilpert (University of Freiburg)

This workshop addresses the methodological theme of the ISLE conference by applying it to the area of diachronic corpus linguistics. In English linguistics, the study of diachronic corpora, i.e. corpora that are divided into sequential time slices, has a long tradition that owes much to the creation of the Helsinki corpus (Kytö 1991). More and more diachronic resources have been made available over the past years, among them ARCHER (Biber 1994), the Penn Parsed Corpora (Kroch et al. 1997), the Corpus of Early English Correspondence (Nurmi et al. 1996), the Corpus of Late Modern English Texts (De Smet 2005), Mark Davies' suite of diachronic corpora (Davies 2007, 2010), and the Old Bailey Corpus (Huber 2007). As a consequence, diachronic corpus linguistics has developed into a topic of considerable methodological and theoretical interest. There now are studies that analyze the diachronic dimension of issues such as frequency effects, collocational interdependencies, or syntactic alternations. Since studies of this kind has only become feasible very recently, many theoretical implications of this work have not yet been fully explored.

The main goal of the present workshop is to bring together researchers working in the field of diachronic corpus linguistics, in order to discuss how state-of-the-art corpus-based methodology serves to enhance the quality of historical description and how it can be used to refine current linguistic theory. The workshop consists of six talks, each of which draws on a diachronic corpus of English to exemplify a modern analytical method, thereby showcasing its strengths and drawbacks and demonstrating how the obtained results go beyond the case study at hand to inform broader questions of linguistic theory. The six conference talks are framed by an introduction and a final discussion.

Participating Abstracts

Introduction: How can new corpus-based techniques advance historical description and linguistic theory?

Hubert Cuyckens (University of Leuven) and Martin Hilpert (University of Freiburg)

In our introduction, we offer a brief survey of different corpus-based approaches to English diachronic linguistics, using the conference theme 'methods - past and current' as a starting point. We then present a number of challenges that future work in diachronic corpus linguistics needs to address and we outline leading questions for the workshop discussion.

Leg it, floor it, snuff it: A synchronic and diachronic analysis of non-referential it

Britta Mondorf (University of Mainz)

While non-referential *it* in subject position has received much scholarly attention (*It is raining*), its use in object position (*to leg it*) has rarely been investigated. When

considering examples such as *leg it*, *floor it*, *snuff it*, we are faced with the question of what motivates the occurrence of *it*.

- (1) Let vs **legge it** a little. (OED: Deacon, J & J. Walker. *Dialog. Disc. Spirits & Divels*. 1601)
- (2) "Follow me!" he yells at Felipe as he swings into his pickup and **floors it**, racing for the stripper as it approaches the end of the route. (COHA Richerson, Carrie. *By Good Intentions*. 2006)
- (3) I have the pleasure to inform you that your mother-in-law **snuffed it**. (OED: *Daily News* 1896)

The present paper presents a corpus-based intervarietal and diachronic study of non-referential *it* with the aim of shedding new light on the role of transitivity in language change. It sets out to investigate whether the occurrence of non-referential *it* can be related to a series of (de-)transitivization processes (cf. e.g. Rohdenburg 2009, Mondorf 2010) that have recently been observed for weakly entrenched verbs or verb senses. Defining transitivity with Hopper & Thompson (1980: 251) as the effectiveness with which an action takes place, the function of *it* might be to enhance a verb's transitivity by equipping it with a pseudo-object. Such semi-transitive uses have also been observed for verbs that do not normally take direct objects, e.g. cognate objects (*to snore a good snore*), verbs occurring in *way*-constructions (*She worked her way to the top*) or with reflexives (*She worked herself to the top*) (cf. Mondorf 2011). The question arises whether the function of non-referential *it* is to modulate transitivity in accordance with the changing entrenchment of the verb with which it is used.

Sociolinguistic variation in morphological productivity in the CEECE

Tanja Säily (University of Helsinki)

According to Adamson (1989: 204), English is remarkable for its “double lexicon”, in which almost all native words have Romance or Latinate synonyms. She argues that this can be seen as a case of diglossia, a sociolinguistic situation in which there is a dual standard of literary and colloquial norms (Adamson 1989: 205–207). The phenomenon is not restricted to the lexicon, however, many derivational affixes of Germanic origin having a synonymous double of French or Latin origin. One of these pairs is formed by the nominal suffixes *-ness* and *-ity*, which have featured prominently in studies of English word-formation. Nevertheless, social aspects of the use of these suffixes have not been addressed satisfactorily, which is surprising considering their diglossic nature. In fact, the question of sociolinguistic variation in morphological productivity in general has received very little attention thus far.

Several measures of morphological productivity have been proposed by Baayen (e.g., 1993), but they are all dependent on the size of the corpus, which makes it difficult to compare measures obtained from sociolinguistically defined subcorpora of different sizes. Säily and Suomela (2009) suggest an assumption-free, highly visual solution based on type accumulation curves and the statistical technique of permutation testing. So far, this method has been applied to 17th-century and present-day materials (Säily and Suomela 2009; Säily forthcoming). Interestingly, both showed that women used *-ity*

significantly less productively than men, while there was no difference in the use of *-ness*. This could imply a gendered discourse style remaining stable throughout the centuries (cf. Nevalainen 2002: 191–194).

The present study analyses 18th-century data from the *Corpus of Early English Correspondence Extension* (CEECE), where initial results seem to show no gender difference with either suffix, disproving the hypothesis of a stable gendered style. Instead, what emerges as significant is an increase in the productivity of *-ity* over time. There also seem to be differences between social ranks, so that the productivity of *-ity* is significantly high among professionals, whereas *-ness* is underused by royalty but used very productively by members of the clergy. However, the differences in the use of *-ness* only appear when the size of the corpus is measured in the number of suffix tokens rather than the number of running words, which raises questions about the theoretical basis of productivity measures. Furthermore, the current method of visualisation seems inadequate, as it can only take into account one measure of corpus size at a time.

Other methodological issues to be discussed include the problem created by exploratory data analysis: the greater the number of observations (in this case, type counts from sociolinguistically defined subcorpora), the higher the likelihood that some of them will be significantly different by chance. This might be solved by applying post-hoc analysis to the results. A further issue is periodisation: if the length of the period under study is changed by a few decades, different factors might emerge as significant, or the significance of some could disappear. Therefore, the current implementation of the method calculates type accumulation curves for several periodisations using a sliding window. Unfortunately, the benefits of the simple visualisation are lost if the user has to browse through hundreds of images of different periodisations. This calls for interactive visualisation (cf. Siirtola et al. forthcoming).

Pairing word order with headedness in the recent history of English: a corpus-based analysis

Javier Perez-Guerra (University of Vigo)

This paper deals with the connection between the type of head and the distribution of its modifiers and complements in noun phrases, adjective phrases and verb phrases. In previous pilot studies, based on Late Modern English data, I showed that the more verbal the head is the more likely the structure of the phrase is to be ruled by principles such as complements-first– rather than by, for example, end weight. On theoretical grounds, the previous finding has consequences for prototypicality considerations affecting verbal, adjectival and nominal heads.

This paper thus focuses on the diachronic tendencies evinced by the data in Early Modern, Late Modern and Present-day English, and pays attention to the structural complexity of the phrasal constituents. The approach is thus corpus-based and analyses data retrieved from different periods and text types in corpora such as the Penn-Helsinki Parsed Corpus of Early Modern English and the Penn Parsed Corpus of Modern British English.

Modeling diachronic change in a morpho-phonemic alternation

Stefan Gries (UC Santa Barbara) and Martin Hilpert (University of Freiburg)

This study addresses the development of the English third person singular suffix from an interdental fricative (*doeth*) to an alveolar fricative (*does*). Based on the Corpus of Early English Correspondence, we analyze more than 20,000 examples from the time between 1417 and 1681 in order to determine (i) in what stages this development took place and (ii) which factors correlate most strongly with each variant.

As for (i), we use a bottom-up clustering method developed specifically for the study of developments over time (Gries and Hilpert 2008 and Gries and Stoll 2009). A multi-step exploratory procedure reveals that the shift from *-(e)th* to *-(e)s* is best characterized as consisting of four to five stages (depending on how one outlier period around 1480 is treated) and as a monotone non-linear increase until *-(e)s* accounts for 90+% of all tokens.

As for (ii), we first coded our data for several factors, including the verb used, whether that verb is lexical (*sing*) or grammatical (*do*, *have*), whether the verb stem ends in a sibilant (*wish*), whether the following word begins with an *s* or *th*, and which variant of the 3SG suffix was used last. Non-linguistic factors included the time period, the identity of the author, and the gender of author and intended recipient.

To predict the chosen variant, we fit a generalized linear mixed-effects model. This model is not only superior to Varbrul analyses (it can include continuous predictors and interactions), but also to standard logistic regressions (it can incorporate verb- and author-specific effects). The resulting model allows us to predict nearly 95% of all inflectional choices correctly and reveals which of the above factors have the strongest influence, thus allowing us to trace the development in a data-driven and highly precise way.

Culture change versus grammar change: the limits of text frequency (and what we can do about it)

Benedikt Szmrecsanyi (University of Freiburg)

This talk is an exercise in variationist model building on the basis of historical corpus data. Any baby – even an otherwise very happy one – will cry when it's getting cold; this doesn't mean that the baby has changed. What has changed is the baby's environment. Historical corpus linguists often face a similar issue, in that fluctuating variant frequencies in real time are a function not only of changing variation grammars, but are also conditioned by changes in discourse norms. Thus, genre-internal fads and fashions (in short: culture change) may affect the textual habitat which, in turn, co-determines variant frequencies via input frequencies. It is precisely because of this that text frequencies as the only gauge for language change may be misleading.

I present quantitative methodologies to keep apart habitat-induced frequency changes and grammar changes proper, tapping the dataset presented in Wolk et al. (forthcoming). Wolk et al. study genitive variability (*the president's speech* versus *the speech of the president*) and dative variability (*John gives Mary a present* versus *John gives a present to Mary*) in ARCHER (A Representative Corpus of Historical English Registers), which covers the period between 1650 and Present-Day English. As will be shown, the input frequency of animacy categories in the corpus is subject to cultural

changes, a fact that inevitably bears on genitive and dative frequencies. I demonstrate how we may use a combination of linear and logistic regression modeling techniques to tease apart culture-induced changes in text frequencies and language change proper.

Corpus-based methodology and grammaticalization theory: Observing, describing, and analyzing grammaticalization and related processes of language change through corpus linguistics

Maria José Lopez-Couso (University of Santiago de Compostela)

The last couple of decades or so have witnessed an increasing awareness of the potential benefits to be derived from the integration of corpus linguistics and grammaticalization theory, two linguistic areas which until quite recently had tended to remain apart. On the one hand, corpus linguistics provides sound empirical methodology for the recognition and documentation of grammatical change; on the other, grammaticalization theory brings corpus linguistics beyond the purely statistical domain. Making use of English diachronic corpora, such as the Helsinki Corpus and ARCHER, this paper illustrates how standard corpus practices can effectively be related to the concerns of grammaticalization theory, thus contributing to our understanding of processes of language change like grammaticalization, subjectification and intersubjectification.

Workshop: Genitive Variation

John Payne (University of Manchester)

Eva Berlage (WWU Münster / University of Paderborn)

The variation between the *s*-genitive construction and other constructions, in particular the oblique *of* construction, has long been a favourite topic of investigation within English linguistics, both from a synchronic and a diachronic perspective. The purpose of this workshop is to explore modern methods of approaching this variation. While the focus of the workshop is on a narrow but important range of constructions, we intend that the methodological discussion will be of interest to a much wider audience. The questions to be explored are:

- 1) What factors are involved in speaker choice between the *s*-genitive and the *of* construction?
- 2) What is the relevance of other constructions, for example the oblique (“double”) genitive and the noun modifier (or compound noun) constructions?
- 3) How has the variation changed over the history of English?
- 4) What methods can be used to investigate these issues?

The primary alternation under investigation in the workshop will be the one illustrated in (1):

- | | | | |
|-----|----|-------------------------|-----------------------|
| (1) | a. | Picasso’s portrait | (<i>s</i> -genitive) |
| | b. | the portrait of Picasso | (<i>of</i> oblique) |

However, two further constructions which have not been examined in the same detail but which also seem relevant are those illustrated in (2):

- | | | | |
|-----|-----|----------------------------|--------------------|
| (2) | a. | this portrait of Picasso’s | (oblique genitive) |
| | b.. | this Picasso portrait | (noun modifier) |

A diverse range of factors have been suggested as forming the basis for the choice between these constructions. Some which are clearly of major significance are:

Syntactic

- definiteness or indefiniteness of the whole construction, as manifested in the choice of determiner
- length/weight of head and dependent nominals

Semantic

- animacy of head and dependent nominals
- semantic relation between head and dependent nominals

However, a host of other factors, including phonological, contextual, pragmatic, and stylistic factors are potentially implicated.

The workshop will explore the potential and difficulties associated with the following methodologies:

(a) *Large-scale multifactorial corpus-based analysis*

Following the pioneering work of Bresnan on the dative alternation, e.g. Bresnan et al (2007), there have been a few recent studies which approach the genitive alternation in a similar manner (e.g. Börjars, Denison & Scott 2009; O'Connor, Maling & Skarabela 2009, Payne & Berlage 2010). Szmrecsanyi (2009) uses a similar approach with a diachronic perspective. Major issues which arise here include the type of coding to be used, the treatment of categorial and near-categorial variation, and the statistical methods.

(b) *Controlled psycholinguistic investigation*

Earlier studies have investigated speaker judgments concerning the genitive alternation (e.g. Rosenbach 2002). What are the difficulties associated with this methodology, and what should we expect when the results of such studies are compared with corpus-based investigations (see in particular Gries 2002)?

(c) *Analysis of examples in context*

Large-scale studies tend to count properties in isolation of context, or at best by operationalising context through some mechanical procedure. What do we lose by using such procedures, and are such losses important?

The most recent workshop on possessive constructions was the workshop on morphosyntactic categories and the expression of possession held at the University of Manchester, 3-4 April, 2009. This had a considerably wider remit than the workshop now proposed. We intend now to invite some of the participants at this workshop, and others, to contribute to a new forum aimed specifically at discussing the possible methodologies used in the investigation of the genitive alternation, and the difficulties associated with each. Of particular concern will be the potential pitfalls of statistically-based approaches.

Participating Abstracts

Genitive variation: the role of the oblique genitive

John Payne (University of Manchester) and Eva Berlage (WWU Münster / University of Paderborn)

In this paper we carry out a corpus-based investigation into the oblique genitive, the construction illustrated by examples such as *some friends of Deborah's*, and identify the factors involved in its selection in modern British English. The investigation builds on that of Payne (to appear), where a narrower set of oblique genitives is studied in isolation.

The oblique genitive tends to be excluded from existing corpus-based or experimental studies of genitive variation, e.g. Rosenbach (2002), O'Connor, Maling & Skarabela (to appear); Börjars, Denison & Scott (to appear), Szmrecsanyi (to appear). Rather, the focus is exclusively on the variation between the *s*-genitive construction, exemplified by *Deborah's friends*, and the *of* construction, exemplified by *the friends of Deborah*. A common methodology is (a) to identify a subset of examples in which

variation between these two constructions is in principle possible, excluding all factors which categorically or near-categorically favour one construction or the other; then (b) to analyse this subset for the effect of factors such as the animacy, length, and discourse status of the dependent. All of these factors turn out to be significant.

The first step in this methodology, narrowing the investigation to examples in which variation between the *s*-genitive and *of* constructions is possible, tends to be operationalised by restricting attention to examples of the *of* construction which are determined by the definite article *the*, and to examples of both constructions in which the dependent consists of more than one (or two) words. This is unfortunately a space from which the oblique genitive, which typically occurs precisely with short dependents and with determiners other than *the*, happens largely to be excluded.

In order to bring the oblique genitive into play, we first consider the semantic relations between head and dependent (strict possession, interpersonal relationship, authorship etc.) which are permitted to the oblique genitive. Fortunately, these represent a sizeable subset of those permitted to the *s*-genitive and the *of* constructions.

Next, we consider a range of syntactic environments in which variation between all three constructions is nevertheless possible. These are typically environments in which the referential anchoring effect of *s*-genitive dependents is weakened and the noun phrase as a whole is in some way non-referential; Rosenbach (2006) provides a detailed discussion of some of these non-canonical *s*-genitive constructions. Consider, for example, noun phrases with *s*-genitive dependents functioning as predicative complements. In the sentence *They were all [Deborah's friends]*, the noun phrase *Deborah's friends* is no longer referential, and the *s*-genitive competes directly both with indefinite occurrences of the *of* construction, as in *They were all [friends of Deborah]*, and the oblique genitive, as in *They were all [friends of Deborah's]*.

By examining such environments, we aim to isolate the factors involved (e.g. nature of dependent, length of dependent, semantic relation between head and dependent), thereby expanding the space in which genitive variation can be quantitatively investigated.

On the genitive's trail: data and method from a sociolinguistic perspective

Sali Tagliamonte and Bridget Jankowski (University of Toronto)

Research on the English genitive (e.g. Rosenbach 2007:154) reports increasing use of the *-s* variant, as in (1-4):

- (1) *A friend's* mother owned it.
- (2) It's not the *horse's* fault when mistakes happen.
- (3) I want to see *Toronto's* home games.
- (4) I think she was washing the *country's* clothes

This has been explained as extension to inanimate possessors, a semantic shift (e.g. Hundt 1997, 1999; Rosenbach 2002), (2-3), and to the pressures of economy in North American journalism, a register change (Hinrichs & Szmrecsanyi 2007; Szmrecsanyi and Hinrichs 2008).

The present work reports on a large-scale sociolinguistic investigation of the genitive in Canadian English using socially stratified corpora and speakers of all ages, but includes only contexts where both forms are grammatical (N=1178). The results show a dramatic split between animates with *-s* (89%) and inanimates with *-of* (92%). Logistic regression and mixed effects modeling of linguistic and social predictors reveal the same suite of conditioning predictors found in studies of written language. Yet when the data are partitioned, we discover that animates, (1-2), are stable over time and use of *-s* correlates with processing constraints, such as length of the possessum and whether the possessor ends in a sibilant. In contrast, inanimates with *-s*, (3-4), are increasing and correlate with possessor length and persistence.

We conclude that the genitive is actually a system in which inanimates have only recently begun a shift towards the *-s* variant. We suggest it is a change from above, driven by external factors such as register but also, word choice and the nature of possessive relationship. In essence, the *-s* genitive is penetrating possessor types that are inanimate, but involve groups of people (countries, cities, sports teams, etc.). We highlight the fact that syntactic alternations of this type are sensitive to grammatical contrasts as well as genre effects. Finally, we emphasize that viable comparisons across studies require a concerted effort towards consistent pre-statistical procedures for data extraction, circumscription of variable contexts and statistical accountability for the highly differentiated inventory of lexical items found across situations.

Is animacy the most important factor in predicting the English possessive alternation?

Cathy O'Connor (Boston University)

Recent corpus-based studies of factors predicting speaker choice of grammatical alternant (e.g. Dative PP vs. double object, Passive vs. Active, and the *s*-possessive vs. *of*-possessive) have revealed robust effects for syntactic weight, information status, and animacy. Arnold et al. (2000), Bresnan et al. (2007), Manning (2003), Rosenbach (2005), Wasow et al. to appear, *inter alia*, have explored the role of these factors in 'soft constraints' and have sought a deeper understanding of their place within linguistic theory.

These studies have brought into focus multivariable statistical methods such as logistic regression. These methods have their own constraints and limitations that determine which data may be included in studies like these. We will argue that for the English possessive alternation, such methodological limitations may have led to a spurious result.

Recent research (e.g. Hinrichs & Szendrői 2007, Rosenbach 2002, 2005) has shown that Jespersen's observations about the "Saxon genitive" or *s*-possessive (e.g. *the child's dream*) versus the "Norman genitive" or *of*-possessive (e.g. *the dream of the child*) are borne out in large-scale studies of corpora as well as in experimental work involving speakers' predictions about which alternant is "better." Holding other factors constant, the studies cited above have all shown that when referring to an *animate* possessor (either human or animal), speakers strongly favor the *s*-possessive (***the teacher's virtues***). When referring to an *inanimate* possessor, speakers strongly favor the *of*-possessive (*the virtues of **walking**; the appearance of **the building***). Similarly, *weight*

figures as an important factor: the longer the possessor phrase is, the more likely the speaker is to choose the *of*-possessive construction (*the victory of **the newest Republican senator*** vs. ***Brown's** victory*). Finally, *discourse status* or *information status* has been shown to be important. Discourse-old possessor referents tend to appear in prenominal position (***his** car*) while discourse-new possessors tend to appear in postnominal position (*the car of **a neighbor across the street***). In these studies, *animacy* consistently emerges as statistically "most important" predictor of usage.

In a study based on over 7,000 examples from the Brown corpus, using logistic regression, we independently found that *animacy* ranked as the strongest predictor of possessive alternant choice. However, we call this result into question for two reasons. First, we will show that the set of NP expressions that most strongly reveal the role of discourse status in predicting use of the *s*-possessive, i.e., *pronouns*, must be discarded from any regression analysis because of the high degree of collinearity they introduce: the factors of weight, animacy and discourse status are massively confounded in pronominal possessors. And while they are not completely ruled out as *of*-possessors (e.g. "*to the left of **him***") they are statistically negligible. Thus the decision to exclude them is methodologically well-founded, yet ineluctably shifts the balance in favor of *animacy* as the strongest factor. And for corpus-based studies using statistical analysis (or experimental designs like Rosenbach's for that matter) there is no way to get around this. Can we say anything more about the relative contributions of animacy and discourse status?

We step back one level and present data based on analysis of texts within three genres found within the Brown corpus (Western fiction, Press, and Non-fiction). The analysis points to text-based features underlying the strong animacy effect, shows they are secondary to discourse status, and provides directions for teasing out the relative contributions of animacy and discourse status in similar studies.

Poss-s vs poss-of revisited

Kersti Börjars, David Denison and Grzegorz Krajewski (University of Manchester)

In this paper we return to the often-studied choice between the two main ways of marking possession in English: possessive 's and an *of*-phrase. Our dataset is the spoken portion of the British National corpus, some 10 million words and over 40,000 possessive NPs, with the data carefully coded and then screened for irrelevant patterns.

We will present a new logistic regression analysis on the choice between *poss-s* and *poss-of*, bringing in such well-studied factors as topicality of possessor and possessum, animacy, final sound of possessor NP, and so on. While confirming the general findings of Rosenbach (2002), Hinrichs & Szmrecsanyi (2007), Szmrecsanyi & Hinrichs (2008) and others, we will show that there are surprising interactions between factors, for which we attempt a linguistic explanation, such as the peculiar status of possessive NPs where the possessum is 1 word long, or those where there is apparent postmodification of the head of the possessor of length 1. Rather than simply looking at overall length of the possessor NP, we distinguish carefully between pre- and postmodification of the head, in line with our previous work on the rarity of the so-called group genitive in naturally-occurring speech (Denison, Scott & Börjars 2010). We show that the presence of postmodification of the head of the possessor NP has a much stronger

effect than the presence of premodification.

Finally, we discuss the theoretical and historical implications of our findings.

Genitive variation in Late Modern English: focus on weight and rhythm

Katharina Ehret, Christoph Wolk, and Benedikt Szmrecsanyi (Freiburg Institute for Advanced Studies)

We investigate the historical development and impact of various factors on the English genitive alternation (*Tom's car* vs. *the car of Tom*) drawing on the richly annotated dataset presented in Wolk et al. (forthcoming). Wolk et al. tap the British news and letters section of ARCHER (A Representative Corpus of Historical English Registers), which covers the period between 1650 and 1990. Wolk et al. then use state-of-the-art multivariate analysis techniques to tease apart the influence of several explanatory factors, such as weight, animacy, definiteness, and semantic relation.

In this talk, we present some of the methodological challenges that were encountered during data annotation and analysis, and extend the research in Wolk et al. by considering further operationalizations of grammatical weight, as well as rhythm as an additional phonological predictor not included in the original Wolk et al. study. Grammatical weight is a well-known factor in syntactic alternations; we focus on the integration of simple single-constituent and more complex multi-constituent measurements. Rhythm – more specifically: the ideally even distribution of stressed and unstressed syllables – is a newcomer to the set of features customarily researched in probabilistic analysis (particularly as far as historical data are concerned). Our take on rhythm is informed by Schlüter's (2005) work on the Principle of Rhythmic Alternation and technically largely follows the method outlined in Shih et al. (submitted).

Pre- and postnominal possessives in English, Dutch and German: An FDG account

Evelien Keizer (University of Amsterdam/ACLC)

The primary aim of this paper will be to find out which factors play a role in determining the choice between a prenominal and a postnominal possessive in English, German and Dutch. The paper will begin with a brief discussion of the use of prenominal and postnominal possessives in English (based on Keizer 2007a; 2007b: Ch. 12). It will be demonstrated that in English there are only very few absolute constraints, the result being that in many cases both constructions are semantically and syntactically acceptable. It will be argued that in those cases a speaker's choice for one of the two constructions is determined by a combination of (morpho-)syntactic, semantic, processing and pragmatic factors, the most important ones being topicality and complexity.

Next, a comparison will be made with similar constructions in German and Dutch, where, it will be shown, the use of the prenominal construction is much more restricted than in English. This is illustrated in (1), where the standard varieties of German and Dutch clearly favour the postnominal construction where English tends to use a prenominal construction:

(1)

<i>English</i>	<i>German</i>	<i>Dutch</i>
The child's book	--	^{*)} Het kinds boek

?? The book of the child	Das Buch des Kindes	Het boek van het kind
--------------------------	---------------------	-----------------------

These restrictions on German and Dutch can plausibly be assumed to be morphosyntactic in nature. Thus, in Dutch NPs, the gender of the head noun determines the form of the article; as this may lead to conflicting forms in the case of a prenominal construction such as given in (1), which contains two consecutive nouns, use of this construction tends to be avoided.

What is interesting, however, is that in certain dialects and/or styles, variation does occur (at least in the case of singular, [+human] possessors):

(2)	<i>German</i>	<i>Dutch</i>
	Dem Kind sein Buch (spoken; informal)	Het kind z'n boek (spoken; informal)
	Des Kindes Buch (poetic; elevated)	

These data, limited as they are, seem to suggest that speakers of German and Dutch still feel the need to create the possibility of choice – a choice which may then be exploited to express additional (in many cases) pragmatic information. Corpus data (ICE-GB, BNC, COCA) will be used to test this hypothesis, and to find out whether, in those cases where variation is possible, the factors determining the choice for either construction are the same as those identified for English.

The paper will end with a brief demonstration of how Functional Discourse Grammar (Hengeveld & Mackenzie 2008) can deal with the interaction between the various (morpho-) syntactic, semantic, processing and pragmatic factors that play a role in the selection of a particular possessive construction.

Workshop: Methods of Analyzing Spoken English

Neal Norrick (University of Saarland)

This workshop brings together an international panel of scholars who study spoken English from diverse perspectives using a range of methods. We will present and justify our own choice of data and our approaches to it, with an eye to identifying complementary aspects of our analyses and discovering areas of intersection.

Among the issues we will tackle from our various standpoints are the composition and range of corpora, the uses of small versus large corpora for qualitative versus quantitative research. Some of the workshop participants have developed their own corpora of spoken English, but all of them have had recourse to a range of corpora for various purposes: American National Corpus (ANC), British National Corpus (BNC), CallHome (Linguistic Data Consortium), ICE GB (International Corpus of British English), Longman Spoken and Written English Corpus, Santa Barbara Corpus of Spoken American English, Wellington Corpus of Spoken New Zealand English. These corpora are annotated and tagged in different ways and offer different methods of data collection. One may begin with a small pre-selected corpus representing a specific set of items or types of interaction to identify a range of functions for a particular linguistic feature before embarking on a general search in a large corpus to determine distributions and frequencies in the corpus at large only to return to a small corpus representing specific contexts for careful qualitative analysis. We will ask whether a bottom-up data-driven approach is useful in the study of discourse particles, interjections and particles, investigating systematic differences between datasets in terms of their collocational patterns, relative frequencies, lexical inventory etc. Panelists will also consider the possibility of analyzing associations between discourse units above the level of the word and the phrase to facilitate a higher level of abstraction and help to uncover the workings of the idiom principle at the discourse level. By contrast, investigations of large corpora may provide the impetus for research, illustrating a range of items or contexts not found in a small corpus and suggesting hypotheses to be tested. For statistical purposes, larger is presumably always better. On the other hand, there is increasing interest in multi-modal data and description in research on spoken language. By its very nature, research on prosody, gaze, gesture and their interrelation within the details of interaction with objects in a specific physical setting must proceed from a narrow recording and transcript of a particular event: Large corpora are extremely difficult to manage in these circumstances. The selection, use and construction of corpora as well as the possibilities and limitations of corpus tagging and annotation represent a major focus of this panel.

Many of the workshop participants maintain that spoken language data must be as authentic as possible, recorded in real life contexts where something is at stake for the participants. But the whole matter of recording, often with camcorders as well as microphones, raises the spectre of the Observer's Paradox. How natural is the language interaction, when we require technical apparatus to record it? Some workshop participants ascribe to the notion that language use must be understood in terms of embedded practices, as part of the social world in which listening and speaking are modalities of action, in line with the treatment of speech production and reception as embodied social facts (Hanks, 1996), and practices as shared ways of doing things, ways of talking, beliefs and values as a function of engagement in an activity (Wenger, 1998;

Eckert, 2000). Seeing talking and listening in terms of practices affords a natural purchase on recurrent and significant features of participant behavior, including gaze and gesture, in face-to-face interaction (Goodwin, 2003). This has consequences for what we count as data, how much context we consider, what we transcribe and in how much detail.

At the same time, certain questions will always go unanswered on the basis of corpus data, no matter how careful the transcription and no matter how large the corpus, since every individual exchange is participant designed and context dependent. Sometimes we are sure we have heard certain items and constructions in everyday talk, but they just do not turn up in any of the corpora we consult. What is the value and reliability of our insights as researchers? We agree that the impressions of naive native speakers count for little in the prosodic analysis of speech. We frequently grapple with the reliability of our auditory perceptions, with or without the support of instrumental analysis. Some studies are built upon close prosodic analysis (boundary placement, accent placement, tone choice) which our readers have to take on trust. And what about data on speech from scripted sources? With some kinds of structures like discourse markers and interjections, there is the problem of how to get enough examples to make sensible generalizations. What other sorts of data are permissible and how trustworthy are they? What value is there in asking the participants themselves about individual interactions? Why should we not tap into the intuitions of native informants by means of questionnaires? Questionnaires offer special advantages as a source of data on speaker attitudes and the social significance of linguistic constructions, especially when triangulated with results from corpus analysis.

Ethnomethodology and CA propose to limit their analyses to categories participants themselves can be seen to orient to. The panel will explore the participants' perspective as a methodological tool in studying prosodic-phonetic details of spoken interaction. But with features such as discourse markers, various syntactic phenomena and prosody as investigated by Interactional Linguistics, it seems as if speakers are not aware of them at all, and they certainly have no terms for the kind of phenomena we are interested in. What kinds of evidence do we require that participants are orienting to particular categories? When are we justified in applying our own external linguistic categories and distinctions?

Discussion of methods in researching spoken language cannot proceed on a theoretical plane alone. Workshop participants will present examples of their data and analyses for discussion. Questions concerning their methods of recording, transcribing and describing talk in interaction will arise naturally as participants exemplify their approaches and arguments on the basis of concrete examples.

Participating Abstracts

Investigating Interjections in Narrative Contexts: A Hybrid Corpus Approach

Neal Norrick (University of Saarland)

In the study of interjections, corpus investigation is particularly important for several reasons: Interjections are not integrated into the grammatical structure of utterances/sentences; they can stand alone as independent units in their own right; they

tend to have functions rather than meanings; they differ by national, regional and personal variety. Corpus research is necessary to reveal their distribution and range of functions. Analysis of interjections in corpora shows them occurring with special types of constructions, e.g. in dialogue within narratives.

Interjections present special problems for corpus research as well. They require careful qualitative analysis because: Interjections as traditionally classified include widely dissimilar items. Common secondary interjections like *boy*, *man*, *hell*, *damn* and *shit* fulfil functions as members of major word classes. Various interjections are familiar in such diverse functions as discourse markers, continuers, attention signals, hesitators, expletives and so on (see e.g. Quirk et al, 1985). They are roughly the same as the *inserts* of Biber et al. (1999), but they are not consistently characterized or marked in corpora. Moreover, the constructions and contexts one determines for interjections on the basis of qualitative investigation are not consistently tagged in large corpora: thus, dialogue is introduced in too many different ways to allow for simple counting in corpora, the various contexts become visible only upon close inspection (compare: *I says gee I ought to get that* versus *laugh and go gee I hadn't seen one* versus *to think gee I survived*).

Further, phrasal interjections and combinations of interjections have been largely ignored in past research, both traditional and corpus-based. Phrasal interjections like *I tell you* can be difficult to distinguish from grammatically integrated phrases without close analysis of spoken data as well. Some combinations of interjections are distinguished only by intonation, so that careful qualitative investigation of spoken data is required for complete analysis (e.g. *oh yeah* with the main stress on *oh* and a falling intonation contour to agree versus *oh yeah* with a rising intonation contour to indicate surprise at the foregoing turn). Consequently, one must create sub-corpora more or less “by hand” on the basis of qualitative analysis, and this conspires with the other factors mentioned to render quantitative corpus results for interjections less robust and convincing than for other parts of speech and phrases.

In my paper, I present the results of a hybrid research approach, employing both large corpora and small, both quantitative corpus studies and qualitative analysis of individual examples. Very large samples of data reveal statistical patterns and variants otherwise hidden, but they must both grow out of small scale qualitative analysis and be tested against it. We must continue to rely on small corpora with their careful transcriptions produced for specific purposes, particularly in cases where prosody plays a role and where constructions are ambiguous, as is often the case with combinations of interjections and with phrasal interjections.

Corpus-driven approaches to discourse markers in spoken data

Gisle Andersen (Norwegian School of Economics and Business Administration)

The paper asks the question: can data-driven approaches be useful for the study of discourse markers in corpora? Corpora have shown to be highly valuable to the study of discourse markers. However, corpus-based studies of discourse markers have almost exclusively applied the method of one-to-one searching (Walsh et al. 2008), where specific linguistic forms are searched for in the corpus. The approach requires the researcher's a priori knowledge of the forms and patterns searched for and their relevance to the study of discourse markers. I wish to explore the idea that the use of discourse

markers may also benefit from a data-driven approach, i.e. bottom-up investigation of systematic differences between datasets in terms of their lexical inventory, collocational patterns, differences in frequency for individual words, etc. There may be good reasons for complementing the method of corpus-based studies with a corpus-driven approach. This has been applied widely and successfully in related fields such as phraseology and lexicography (e.g. Sinclair 1991; Partington 1998; Biber 2009). As Aijmer (2002: 31) observes “[a]mong the characteristic properties of discourse particles which have not received enough attention in the literature is their ‘clustering’ (collocations with other particles)”. I am proposing that a corpus-driven approach may be valuable for the identification of new or emergent discourse markers and for changes in the functional range of existing markers. New forms may represent marker usage (e.g. *duh*), a rapid increase in frequency of an item may be due to innovative marker usage (e.g. *you get me*); and changes in collocational patterns may be the result of reanalysis of items taking on discourse functions (e.g. the irony marker *yeah right* and the (sometimes ironical) surprise marker *no shit* (Andersen 2006; Norrick Forthcoming 2012). The paper reports on a series of experiments based on English spoken corpora, such as COLT, recorded in the early 1990s and more recent spoken corpora such as the Linguistic Innovators Corpus (LIC). I investigate monomorphemic and phrasal items through various computational techniques for extraction of frequency data, neology identification, collocation analysis and identification of multi-word units in corpus data, and evaluate the relevance of these methods for the study of discourse markers.

The participants' perspective: Interactional-linguistic work on the phonetics of talk-in-interaction

Dagmar Barth-Weingarten (University of Freiburg)

The so-called "participants' perspective" is a central methodological principle in conversation-analytic (CA) work. CA aims at discovering the organisation and order of social action in interaction (Psathas 1995). It assumes not only that everyday talk is systematically organized and deeply ordered at all points (e.g., Heritage 1984), but that this orderliness is displayed and used by the participants to accomplish actions and activities for, and with, each other given the sequential, and other, contingencies of talk (cf. Schegloff/Sacks 1973: 290, Hutchby/Wooffitt 1998). At the same time, the analyst can observe this orderliness to infer how the participants do so from an emic viewpoint, i.e. from within the system (Pike 1967), namely the participants own handling and understanding of the system (Seedhouse 2005). CA practitioners investigate individual instances of talk in terms of what is being accomplished, how this is done and how it is oriented towards by the participants in order to abstract from that the underlying system in a data-driven way. Once described, further evidence for the working of the system needs to be provided by analyses of further instances as well as the incorporation of deviant cases along the same lines (Wootton 1989).

Interactional Linguistics (IL) has adopted these assumptions and methods from CA and seeks to study how the participants employ the linguistic details of talk to jointly accomplish interaction, including drawing inferences about what is going on and what is required from them at a particular point in interaction (Couper-Kuhlen/Selting 1996, 2001). Together with the York Phonology for Conversation (cf., e.g., Local/Kelly/Wells

1986, Kelly/Local 1989, Local/Walker 2005), it also reconstructs prosodic-phonetic features as member's devices in the organization and management of talk (cf. Couper-Kuhlen/Ford 2004, Barth-Weingarten/Reber/Selting 2010, for instance).

Along these lines, this presentation seeks to highlight the participants' perspective as a methodological tool in studying prosodic-phonetic details of spoken interaction. By means of excerpts from American-English telephone conversations from the CallHome corpus, it discusses what this methodological tool can, and cannot, provide us as analysts with in these regards. In particular, it will focus on the use of the participants' perspective in investigating phonetic cues to 'ceasuring', i.e. structuring (Auer 2010), of talk. Its advantages include providing us with the foundations of a new approach to what scholars have so far – with particularly little success for natural everyday talk (Cruttenden 1997, for instance) – attempted to capture by the concepts of "intonation units", "declination units", "spoken paragraphs" etc. The alternative suggestion is in line with the aim to postulate categories that can be shown to be relevant for the participants and to be oriented towards by them. At the same time, the participants' perspective has a number of implications for analysis, including the actual method of phonetic analysis (natural data, auditory analysis, audible cues, for instance), while it also has clear boundaries for instance in terms of the actual phonetic analysis itself (participants cannot describe, let alone exhaustively list, relevant phonetic cues). Yet, the latter can be circumvented exactly by focusing on how the participants orient towards them.

Studying DM Sequences in Spoken English

Bruce Fraser (Boston University)

The problem I am addressing involves sequences of DMs. I have in mind sequences like the following (where the sequence is in **bold**):

1. John didn't go right home. **But instead**, he went first to a bar.
2. Mary doesn't like blackberries. **Rather, on the other hand**, she likes currents.
3. Bill got very angry. **Thus, as a result**, he left early.
4. We went on a picnic. There were ants **and, in addition**, it rained.

There is an unspoken assumption, among some, that if a sequence is found in COCA or BNC (even though these corpora are essentially of written language), that the DM sequences are acceptable in spoken English. However, a survey of native English speakers shows that depending on who is speaker and the utterance context, there is considerable difference of opinion as to the acceptability of the sequences. Some are clearly acceptable, some clearly unacceptable, and some in the middle.

I will present the results of two experimental protocols that take recordings of such sequences in conversations and have native speakers judge on a scale of 1-5 the acceptability, both when they are themselves the speakers and when they are listeners. In addition, I will suggest an explanation for questionable acceptability.

Introducing collocation analysis

Christoph Rühlemann (University of Munich)

Much progress has been made in recent years in the study of what Sinclair (1991) referred to as the ‘idiom principle’: the co-selection of words at the level of phraseology. A sizable number of building blocks of the idiom principle have come under scrutiny, including n-grams, conc-grams, collocation, colligation, collostruction, semantic preference, semantic prosody, and textual colligation. These building blocks of syntagmatic patterning operate on different levels of abstraction from linear surface patterning. At the most surface-bound and, thus, lowest level of abstraction we find n-grams, which can be described as contiguous word sequences. At the highest level of abstraction from linear surface structure we find textual colligation, which is an extension of the notion of colligation in the sense that textual colligation describes the association that holds between a lexical item and a particular position in a text (Hoey 2005). Located between these extremes we find, in ascending order of abstraction, concgrams, collocation, colligation, collostruction, semantic preference, and semantic prosody. While these building blocks of the idiom principle can be distinguished in terms of abstraction from purely linear patterning, they crucially share one property: the lexical item or sequence of lexical items. As a result, methods used to investigate these association types are necessarily lexis-driven. However, given that discourse essentially operates on levels above surface structure, lexis-driven analyses of discourse are not always sufficient to ‘crack’ discourse.

In this paper I introduce a new quantitative method I will refer to as ‘collocation analysis’, a term by which I mean the analysis of associations involving discourse units (the Greek word ‘logos’ also meaning ‘discourse’). This method facilitates a still higher level of abstraction from lexical surface structure and can help uncover the workings of the idiom principle at discourse level. Collocation analysis can be carried out in two basic ways: (i) as an extension of the existing lexis-driven methods and (ii) independently of lexically driven methods. Variant (i) uses a lexis-driven analysis as its starting point and gradually abstracts away from it to higher lexis-independent discourse levels; variant (ii) investigates the co-selection of lexis-independent discourse units directly.

The requirements for quantitative collocation analysis are twofold. First, collocation analysis requires discourse-annotated corpora. The corpus I will base this paper on is the Narrative Corpus, a 150,000-word corpus containing conversational stories embedded in their conversational contexts. The texts in the corpus have been extracted from the conversational subcorpus of the British National Corpus and contain all the original POS and XML annotation. Additionally, the NC has been consistently annotated for a number of discourse features, including textual components, participant status, quotatives, and reporting modes. Second, another prerequisite for implementing collocation analysis is the use of appropriate corpus-analytical tools. The tool I will be briefly commenting on is XPath, a query language which operates on the hierarchical tree structure of XML texts.

To illustrate collocation analysis I will outline a case study which aims to test the hypothesis that narrators use stimuli to encourage recipients to respond (cf. Norrick 2008). Overall, I wish to demonstrate the usefulness of working with richly discourse-annotated corpora, which lies in their potential to reveal associations far above lexical surface structure.

Just how useless are questionnaires for studying spoken language?: Triangulating elicited and natural corpus data

Klaus P. Schneider (University of Bonn)

‘Questionnaire-bashing’ is extremely popular among scholars interested in spoken language (cf., e.g., Golato 2005). What seems to be forgotten in this discussion are two of the most fundamental methodological ‘laws’. These are: (1) There is no best method. Each and every method of data collection has advantages as well as disadvantages. (2) Choice of method depends entirely on the research questions to be answered (cf. Kasper 2008). This means that even naturally occurring discourse, i.e. the only data type accepted in some circles (especially in conversation analysis and interactional linguistics), is not per se superior and is not suitable for all research purposes.

To balance out the respective weaknesses of individual methods, different types of data should be triangulated (cf., e.g., Dörnyei 2007). In the present paper a case is made for combining data from electronic corpora of naturally occurring spoken language with experimental data elicited by employing production questionnaires. Using questionnaires helps to overcome the well-known problems of precision and recall which occur when using large electronic corpora in empirical pragmatics (cf. Jucker 2009). Such corpora can be easily searched for discourse markers and similar phenomena, i.e. when the starting point is a form or position and not a function or illocution (cf. O’Keeffe et al. 2011). In the latter case, e.g. for studies of speech act realization or sequential patterns, manual search is inevitable as no pragmatically annotated corpus is yet available (cf. Kohnen 2008). Automated corpus search is, however, also possible if based on realization forms elicited in questionnaires. In other words, questionnaires can be used to generate realizations, and corpora can be used to determine their status in terms of frequency and distribution.

A further advantage of using production questionnaires is that questionnaire data are directly comparable (cf. Schneider 2010b). Comparability is not easily achieved when using corpora of naturally occurring spoken language. In this case, situations may differ significantly in terms of time, place, occasion, activity type and discourse genre and, more importantly, regarding the identities of the participants and their relationships. Often, only limited information is available on these parameters, if at all. In questionnaires, on the other hand, all relevant variables are known and, more crucially, can be manipulated in systematic ways. Therefore, questionnaire data are obviously experimental data, even though they are not collected by employing genuinely psycholinguistic methods (cf. Noveck & Sperber 2004).

Finally, the nature and limitations of questionnaire data have to be born in mind. Needless to say, such data do not represent what participants actually do in spoken interaction, but what they think they would do or should do in a given type of social situation. Put another way, questionnaire data reflect what is expected and considered appropriate and can, hence, be regarded as the pragmatic norm or convention. Thus, such data reveal scripted behaviour, i.e. the prototypes and cognitive schemata underlying actual behaviour (cf. Schneider 2010a). In this sense, questionnaire data can be seen as competence data (cf. Schneider 2011). Triangulating questionnaire data with corpus data makes it possible to compare the norms to actual performance.

The issues addressed here are illustrated with examples from on-going projects in variational pragmatics, in which discourse completion tasks and dialogue production tasks are combined with material from various components of the International Corpus of English, the Santa Barbara Corpus of Spoken American English, and the spoken part of the Corpus of Contemporary American English as well as a small self-recorded corpus of Irish English.

Inconsistencies in the prosodic analysis of corpus data: evidence from grammaticalisation

Anne Wichmann (University of Central Lancashire) and Nicole Dehé (University of Konstanz)

Existing spoken corpora are a rich source of data, and some annotations (e.g. part of speech tagging) are relatively easy to achieve once an orthographic transcription is available. In some cases, the compilers have not made the original sound recordings freely available (e.g. British National Corpus), and the transcription must then be treated as the primary data. Those interested in prosodic analysis need either a prosodically annotated corpus (e.g. London Lund Corpus, Santa Barbara Corpus of Spoken American English) or access to the sound files. Prosodic annotation is not, however, straightforward. The context of some recordings – speech in its natural habitat – means that their quality does not permit reliable instrumental analysis. An auditory analysis is a time-consuming business, and is generally left to the consumer. As a result, prosodic research into naturally-occurring speech is based heavily on the analysis of individual researchers. In some cases, these analyses are restricted to those features that can be extracted automatically (e.g. pitch range, speech rate) (Shriberg et al 1998), but linguistically motivated analyses cannot be achieved by automatic methods alone. Many published papers are based on a mixture of auditory and instrumental analysis, and readers have no choice but to trust the author's analysis. But even highly trained prosodic labellers do not always agree (Knowles et al 1996), and this means that the annotation on which such papers are based may be open to challenge. This concerns prosodic features such as tonal and temporal cues to prosodic boundaries, identification of nuclear v non-nuclear accents and their phonetic manifestation, the interpretation of tonal movement, and the like. We have investigated some of these cases and find a worrying inconsistency that undermines the claims being made. For example, accents relate to prominence and prominence relates to meaning. Similarly, prosodic phrasing may resolve semantic disambiguities. Inconsistencies in prosodic labelling may thus affect prosodic theorising and theoretical approaches to the prosody-meaning relation.

As an example, we will present some of our recent work on expressions in English such as the sentence adverbial *of course* (Wichmann et al 2010) and comment clauses (*I think, I suppose* etc) (Dehé & Wichmann 2010a, 2010b). With these items, we observe an ongoing process of grammaticalisation (or pragmaticalisation) through evidential meaning to interpersonal meaning. In other words, these expressions have acquired in some environments the status of pragmatic markers. The prosodic evidence for these differences in usage and interpretation crucially relies on accurate identification of prosodic boundaries and prominences. For example, whether sentence-initial *I think* or *I believe* functions as main clause, comment clause or discourse marker is reflected in accent placement: in particular, on the presence or accent of an accent and on its exact position.

By examining cases like these, and pointing out the nature of the inconsistencies we have found, we wish to raise awareness of some of the problems associated with prosodic analysis.

Transportation

For taxi service, call Boston Cab at 617-536-5010. It is also generally easy to hail a cab on the street.

Taxis are widely available throughout Boston, but the city's public transit system (the MBTA) is convenient and relatively simple to navigate. MBTA fares are \$2.00 per train ride and \$1.50 per bus ride using cash or a paper ticket. Discounted rates (\$1.70/train, \$1.25/bus) are available with a plastic "Charlie Card." These can be obtained at most MBTA train stops. Ask an operator for a card, which can then be loaded with the value of your choice at one of the machines outside the turnstiles. Alternatively, a 7-day paper pass can be purchased at the machines for \$15.

Boston University and many local hotels and restaurants are located along the "B" branch of the Green Line; you can use the MBTA's online trip planner to find the best stop for your destination. For direct outbound service to Boston University, be sure to wait for a "B" train (the outbound Green Line splits into four branches) and exit at one of the BU stops. BU East and BU Central are most convenient to the conference; BU West is closest to campus housing and the fitness center.

Shuttle bus to ISLE's extracurricular events (Monday's harbor cruise and Tuesday's walking tour) will depart from outside 725 Commonwealth Avenue (CAS).

Campus Facilities

Internet Access

ISLE guests have access to BU's public computer terminals and wireless network. A BU login name and password are required to access the internet. For the duration of the conference, we have set up the following account, which all are welcome to use:

Login name: islenet
Password: Isle2Boston

Fitness and Recreation

BU's fitness and recreation center (FitRec) is located at 915 Commonwealth Avenue. ISLE conference participants are welcome to use the center. A 1-week membership can be purchased for \$40, and a one-day pass to use the gym is \$10, cash only. Memberships can be paid by cash, check, or credit card and can be purchased at the front desk.

Eating Near BU

Lunch and Dinner Suggestions

Sunset Cantina –casual; Mexican/American
916 Commonwealth Avenue, Boston
(617) 731-8646

Noodle Street –causal; Thai
627 Commonwealth Avenue, Boston
(617) 536-1300

Audubon Circle Restaurant/Bar –casual; American/cocktails
838 Beacon Street, Boston
(617) 421-1910

The Elephant Walk –French-Cambodian bistro
900 Beacon Street, Boston
(617) 247-1500

Eastern Standard –American bistro; indoor/outdoor dining; cocktails
528 Commonwealth Avenue, Boston
(617) 532-9100

Cornwall's –causal; pub
627 Commonwealth Avenue, Boston
(617) 262-3749

Kavuga –causal; Japanese
1030 Commonwealth Avenue, Boston
(617) 566-8888

Upper Crust –takeout; pizza
888 Commonwealth Avenue, Brookline
(617) 731-3000

Breakfast and Lunch Suggestions

Panera Bread –open 6am M-F; 6.30 am Sat/Sun
888 Commonwealth Avenue, Brookline
(617) 738-1501

Angora Café –open 8.30am M-F; 9.30 am Sat/Sun
1024 Commonwealth Avenue, Boston
(617) 232-1757

Jamba Juice: open 10am M-F
775 Commonwealth Avenue, Boston
(617) 353-2991

Espresso Royale –open 7am M-F; 8am Sat/Sun
736 Commonwealth Avenue, Boston
(617) 277-8737

Starbucks –varied hours
775 Commonwealth Avenue, Boston –open 9am M-F
(617) 353-2991
847 Commonwealth Avenue, Boston –open 6.15am; Sat 7am; Sun 7.30am
(617) 734-3691

Japonaise Bakery –open 8am all days
1020 Beacon Street, Brookline
(617) 566-7730

Bruegger's Bagels –open 6am all days
644 Commonwealth Avenue, Boston
(617) 262-7939

Boston Activities

Museum of Fine Arts, Boston

<http://www.mfa.org/>

Founded in 1870, Boston's Museum of Fine Arts is one of the largest museums in the United States, containing over 450,000 works of art. Some highlights of the MFA's collection include:

- Egyptian artifacts including sculptures, sarcophagi, and jewelry
- French impressionist and post-impressionist works
- 18th and 19th century American art, including many works by John Singleton Copley, Winslow Homer and John Singer Sargent
- Extensive collection of Chinese painting, calligraphy and imperial Chinese art
- The largest museum collection of Japanese works under one roof in the world outside of Japan

Isabella Stewart Gardner Museum

<http://www.gardnermuseum.org/>

The Gardner Museum has remained virtually unchanged since its founder's death in 1924. Three floors of galleries surround a garden courtyard blooming with life in all seasons. The galleries are filled with paintings, sculpture, tapestries, furniture, and decorative arts from cultures spanning thirty centuries.

Boston's Freedom Trail

<http://www.cityofboston.gov/freedomtrail/>

The Freedom Trail is a red (mostly brick) path through downtown Boston that leads to 16 significant historic sites. It is a 2.5-mile self-guided walk from Boston Common to Bunker Hill Monument in Charlestown with simple ground markers, graveyards, notable churches and other buildings, and a historic naval frigate along the way. Most sites are free; three, the Old South Meeting House, the Old State House, and the Paul Revere House, have small admission fees, while others suggest donations.

New England Aquarium

<http://www.neaq.org/>

With a mission "to present, promote and protect the world of water," the New England Aquarium remains one of the few such institutions with commitments to research and conservation as well as education and entertainment. In addition to the main aquarium building, attractions include the Simons IMAX Theatre and the New England Aquarium Whale Watch, which operates from April through November.

Boston Public Garden

http://www.cityofboston.gov/parks/emerald/public_garden.asp

Together with the Boston Common, these two parks form the northern terminus of the Emerald Necklace, a long string of parks designed by Frederick Law Olmsted. While the Common is primarily unstructured open space, the Public Garden contains a lake and a large series of formal plantings that are maintained by the city and others and vary from

season to season. During the warmer seasons, the 4-acre lagoon in the Garden is usually the home of one or more swans and is always the site of the Swan Boats, a famous Boston tourist attraction dating from 1877.

Newbury Street, Boston

<http://www.newbury-st.com/>

East of Massachusetts Avenue, Newbury Street is lined with historic 19th-century brownstones that contain hundreds of shops and restaurants, making it a popular destination for tourists and locals. The most expensive boutiques are located near the Boston Public Garden end of Newbury Street. The shops gradually become slightly less expensive and more bohemian toward Massachusetts Avenue.

Harvard Square, Cambridge

<http://www.harvardsquare.com/>

At the historic center of Cambridge (northbound from Boston on the MBTA Red Line), and adjacent to Harvard University, the Square features numerous shops, restaurants, and historic monuments. It is also home to many new and used bookstores, including Schoenhof's, a mecca for books on languages and linguistics.

Arnold Arboretum

<http://arboretum.harvard.edu/>

The Arnold Arboretum, maintained by Harvard University and located south of Boston down the MBTA Green Line, is the oldest public arboretum in North America and one of the world's leading centers for the study of plants and trees. A link in Boston's Emerald Necklace of parks, the Arboretum is a unique blend of respected research institution and beautiful public landscape.

Contact Information

Should any questions or difficulties arise during the conference, please don't hesitate to call or e-mail a conference organizer or assistant:

Eugene Green	857-257-5915	eugreen@bu.edu
Bruce Fraser	617-353-3234	bfraser@bu.edu
Emily Griffiths Jones	703-283-9560	emilvg@bu.edu
Karen Guendel	781-296-6150	kareng@bu.edu

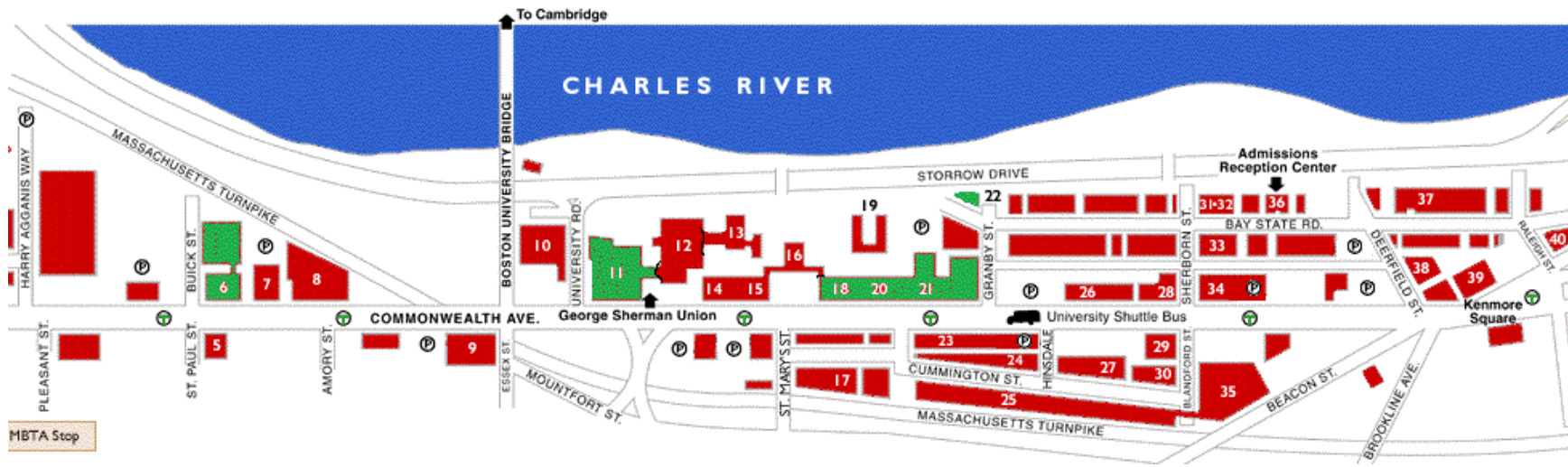
ISLE Volunteers are also available to assist you in CAS 220 during conference hours.

Other useful numbers:

Boston Cab	617-536-5010
BU Campus Security	888-255-9411
BU Buildings & Grounds	617-777-1100
BU Media Support*	617-201-8434 (weekdays)
	617-201-8432 (Saturday and Sunday)

*Please note that this number is provided for conference-related issues with the classroom projector systems, not for software problems or individual help.

BU Campus Map



Buildings in GREEN are relevant to ISLE:

- Building 6: BU housing at 10 Buick Street
- Building 11: George Sherman Union (ISLE Sponsored Dinner in the 5th floor Faculty Dining Room)
- Building 18-21: College of Arts and Sciences (CAS): ISLE proceedings here
- behind 21, off Storow Drive: BU Castle (Pub for opening night reception located at Granby Street entrance on lower level)

Parking lots are marked with a circled P; MBTA stops are marked with a circled T.

Most restaurants listed on pages 147-148, among a number of others, are located along Commonwealth Avenue or Beacon Street.

Index

A

Aarts, Bas	36
Alam, Farhana	110
Andersen, Gisle	140-141
Aronoff, Mark	93-94
Axelsson, Karin	37

B

Bakos, Jon	81
Bao, Zhiming	124-125
Baranowski, Maciej	100-101
Barth-Weingarten, Dagmar	141-142
Bergs, Alexander	57
Berlage, Eva	27, 131-133
Bickerstaff, Garrison	84
Bieswanger, Markus	54
Blanco-Suárez, Zeltia	91
Boberg, Charles	64
Börjars, Kersti	135-136
Bowie, Jill	36
Brems, Lieselotte	40
Bresnan, Joan	28
Brinton, Laurel	78

C

Clark, Lynn	98-100, 101-102
Chapman, Don	51
Cuyckens, Hubert	126
Czerniak, Izabela	41

D

Davydova, Julia	66, 74
Dehé, Nicole	145
Denison, David	21, 135-136
Depraetere, Ilse	34
Deuber, Dagmar	107
Diemer, Stefan	89
Dollinger, Stefan	65, 112
Drager, Katie	102-103

E

Eddington, David	69
Ehret, Katharina	136
Elenbaas, Marion	31
Erlewine, Michael Yoshitaka	90

F

Filppula, Markku	119-121
Franceschi, Daniele	92

Fraser, Bruce	141
Fuchs, Robert	60
<u>G</u>	
Garretson, Gregory	88
Geeraerts, Dirk	46
Green, Lisa	26
Gries, Stefan	29, 128-129
Grue, Dustin	75
Günther, Christine	55
Gut, Ulrike	60
<u>H</u>	
Hakala, Henrik	121-122
Hall-Lew, Lauren	103
Hampe, Beate	79
Hay, Jennifer	102-103
Hazen, Kirk	52, 113
Hilpert, Martin	126, 128-129
Hinrichs, Lars	58, 112
Hoffmann, Thomas	42
Horn, Laurence	25
Huber, Magnus	59
Hundt, Marianne	106, 107-108
Hynninen, Niina	121-122
<u>I</u>	
Ifukor, Presley	72-73
<u>J</u>	
Jankowski, Bridget	133-134
Jansen, Sandra	53
<u>K</u>	
Kaltenböck, Gunther	43
Keizer, Evelien	136-137
Kerremans, Daphné	82
Klemola, Juhani	119-121
Krajewski, Grzegorz	135-136
Kretzschmar, Bill	113
Kurtes, Svetlana	61
<u>L</u>	
Langford, Chad	34
Larina, Tatiana	61
Lazar, Izabela	115-117
Leimgruber, Jakob	95, 108
Leung, Glenda Alicia	107
Lindsay, Mark	93-94
Lopez-Couso, Maria José	130
Lucek, Stephen	68

M

Maguire, Warren	98-100
Maier, Georg	74, 80
Mair, Christian	20
Mauranen, Anna	119-121
McMahon, April	22
Meriläinen, Lea	122
Mesthrie, Rajend	108-109, 125
Micciulla, Linnea	38
Miura, Ayumi	39
Mondorf, Britta	126-127
Montgomery, Chris	50

N

Nesselhauf, Nadja	96
Nevalainen, Terttu	85
Norrick, Neal	138-140
Nycz, Jennifer	103-104
Nykiel, Joanna	118

O

Oda, Toshihiro	71
O'Connor, Cathy	134-135
O'Donnell, Matthew	86

P

Palander-Collin, Minna	76
Parviainen, Hanna	123-124
Paulasto, Heli	122
Payne, John	27, 131-133
Perez-Guerra, Javier	128
Pfaff, Meike	56
Prado-Alonso, Carlos	47
Purnell, Thomas	113

R

Raimy, Eric	113
Ranta, Elina	122
Rathore, Claudia	109-110
Reed, Marnie	114
Ricks, Christopher	23
Rosenbach, Anette	28
Rühlemann, Christoph	142-143

S

Sailer, Manfred	62
Säily, Tanja	85, 127-128
Salmons, Joseph	113
Sanchez-Stockhammer, Christina	63
Sankaran, Lavanya	95

Schmid, Hans-Jörg	82
Schmidt, Sebastian	59
Schneider, Edgar	122-123
Schneider, Klaus	144
Schönefeld, Doris	35
Schulze, Rainer	48
Shank, Christopher	30
Siemund, Peter	74, 121
Siirtola, Harri	85
Speelman, Dirk	46
Stegmayr, Susanne	82
Suhr, Carla	97
Suryanarayan, Neelakshi	61
Stuart-Smith, Jane	110
Stvan, Laurel Smith	49
Szmrecsanyi, Benedikt	28, 129-130, 136
<u>T</u>	
Tagliamonte, Sali	24, 133-134
Tipton, Phillip	104-105
Torrado, Úrsula Kirsten	83
Trousdale, Graeme	33
<u>V</u>	
Van Bogaert, Julie	30
Vandelanotte, Lieven	87
van Hattum, Marije	67
Vartiainen, Turo	44
<u>W</u>	
Wallis, Sean	36
Walker, Jim	77
Wärnsby, Anna	45
Watson, Kevin	98-100, 101-102
Weckwerth, Jaroslaw	114
Wichmann, Anne	145
Wiltshire, Caroline	70
Wolfram, Walt	112
Wolk, Christoph	28, 136
Wulff, Stefanie	29, 32
<u>Z</u>	
Zenner, Eline	46
Zipp, Lena	110-111