Corpus Approaches to Contemporary British Speech

Featuring contributions from an international team of leading and up-and-coming scholars, this innovative volume provides a comprehensive sociolinguistic picture of current spoken British English based on the Spoken BNC2014, a brand new corpus of British speech. The book begins with short introductions highlighting the state of the art in three major areas of corpus-based sociolinguistics, while the remaining chapters feature empirical studies based on the Spoken BNC2014 data. These analyses focus on English spoken in everyday situations in the UK, with brief summaries reflecting on the sociolinguistic implications of this research included at the end of each chapter. The Spoken BNC2014 as a robust dataset allows this team of researchers the unique opportunity to focus on speaker characteristics such as gender, age, dialect and socio-economic status, to examine a range of sociolinguistic dimensions, including grammar, pragmatics and discourse, and to reflect on the major changes that have occurred in British society since the last corpus was compiled in the 1990s. This dynamic new contribution to the burgeoning field of corpus-based sociolinguistics is key reading for students and scholars in sociolinguistics, corpus linguistics, pragmatics, grammar and British English.

Vaclav Brezina is a Lecturer at the Department of Linguistics and English Language, Lancaster University. His research interests are in the areas of corpus design and methodology, sociolinguistics and statistics. He is an author of Statistics in Corpus Linguistics: A Practical Guide (CUP, 2018). He also designed a number of different tools for corpus analysis such as #LancsBox, BNC64, Lancaster Vocab Tool and Lancaster Stats Tool online. He has been involved in the development of corpora such as the Spoken BNC2014, Trinity Lancaster Corpus and Guangwai-Lancaster Corpus.

Robbie Love is a Research Fellow at the School of Education, University of Leeds, with research interests in applied and corpus linguistics. He completed his PhD at Lancaster University in 2017, where he was lead researcher in the development of the Spoken BNC2014. Before moving to Leeds, he held a post-doctoral position at Cambridge Assessment English, where he worked on the development of the Cambridge Learner Corpus.

15 **Spoken Corpus Linguistics**
From Monomodal to Multimodal
*Svenja Adolphs and Ronald Carter*

16 **Digital Literary Studies**
Corpus Approaches to Poetry, Prose, and Drama
*David L. Hoover, Jonathan Culpeper, and Kieran O’Halloran*

17 **Triangulating Methodological Approaches in Corpus Linguistics Research**
*Edited by Paul Baker and Jesse Egbert*

18 **The Language of Hate**
A Corpus Linguistic Analysis of White Supremacist Language
*Andrew Brindle*

19 **Metaphor, Cancer and the End of Life**
A Corpus-Based Study
*Elena Semino, Zsófia Demjén, Andrew Hardie, Sheila Payne and Paul Rayson*

19 **Understanding Metaphor through Corpora**
A Case Study of Metaphors in Nineteenth Century Writing
*Katie J. Patterson*

20 **TESOL Student Teacher Discourse**
A Corpus-Based Analysis of Online and Face-to-Face Interactions
*Elaine Riordan*

21 **Corpus Approaches to Contemporary British Speech**
Sociolinguistic Studies of the Spoken BNC2014
*Edited by Vaclav Brezina, Robbie Love and Karin Aijmer*

For a full list of titles in this series, please visit www.routledge.com
Corpus Approaches to Contemporary British Speech
Sociolinguistic Studies of the Spoken BNC2014

Edited by
Vaclav Brezina, Robbie Love
and Karin Aijmer
Contents

Preface vii
Acknowledgements viii

PART I
Short Introductions to Corpus-Based Sociolinguistics and the BNC2014 1

1 Corpus Linguistics and Sociolinguistics: Introducing the Spoken BNC2014 3
   VACLAV BREZINA, ROBBIE LOVE AND KARIN AIJMER

2 The Spoken BNC2014: The Corpus Linguistic Perspective 10
   TONY MCENERY

3 Current British English: The Sociolinguistic Perspective 16
   BEATRIX BUSSE

4 Using the Spoken BNC2014 in CQPweb 27
   ANDREW HARDIE

PART II
Discourse, Pragmatics and Interaction 31

5 Politeness Variation in England: A North-South Divide? 33
   JONATHAN CULPEPER AND MATHEW GILLINGS

6 ‘That’s Well Bad’: Some New Intensifiers in Spoken British English 60
   KARIN AIJMER
vi Contents

7 Canonical Tag Questions in Contemporary British English 96
   KARIN AXELSSON

8 Yeah, Yeah Yeah or Yeah No That’s Right: A Multifactorial Analysis of the Selection of Backchannel Structures in British English 120
   DEANNA WONG AND HAIDEE KRUGER

PART III
Morphosyntax 157

9 Variation in the Productivity of Adjective Comparison in Present-Day English 159
   TANJA SÄILY, VICTORINA GONZÁLEZ-DÍAZ AND JUKKA SUOMELA

10 The Dative Alternation Revisited: Fresh Insights From Contemporary British Spoken Data 185
   GARD B. JENSET, BARBARA MCGILLIVRAY AND MICHAEL RUNDELL

11 ‘You Still Talking to Me?’: The Zero Auxiliary Progressive in Spoken British English Twenty Years On 209
   ANDREW CAINES, MICHAEL MCCARTHY AND PAULA BUTTERY

12 ‘You Can Just Give Those Documents to Myself’: Untriggered Reflexive Pronouns in 21st-Century Spoken British English 235
   LAURA L. PATERSON

Contributors 256
Index 261
Preface

The volume is a contribution to the area of corpus-based sociolinguistics. Its main aim is to provide new insights into current spoken British English based on the Spoken British National Corpus 2014 (Spoken BNC2014), a brand new corpus of British speech (collected around 2014). The Spoken BNC2014 is a subset of the new British National Corpus (BNC2014), which is currently being compiled at Lancaster University. The Spoken BNC2014 samples English used in everyday situations across the UK, with a particular focus on speaker characteristics such as gender, age, region and socio-economic status. It therefore allows an in-depth exploration of social factors that play a crucial role in the use of language in current British society.

The idea for this book started in December 2015, when we received an overwhelmingly positive response to our call for applications for early data access to the Spoken BNC2014. The Spoken BNC2014 Sample was made available to the authors of twelve successful proposals. These were selected based on their innovative use of the data and the significance of the topic. Eight of the resulting research papers are featured in the current volume, while the remaining ones can be found in a special issue of the International Journal of Corpus Linguistics (IJCL, 22(3), 2017). The publication of both this volume and the special issue of IJCL are intended to celebrate the public launch of the Spoken BNC2014 in September 2017 and demonstrate its potential for research on current British speech. Our hope is that the readers will find the exploration of current British speech and the changes that have taken place over the period of the last twenty years as exciting as we did when we were reading through the chapters and preparing the volume.
We wish to thank the authors of the chapters for their contributions bringing unique insight into current British speech. The chapters in this volume were externally reviewed; we are greatly indebted to the external reviewers for their time and helpful comments on the submissions. We would like to thank Mercedes Durham, Signe Oksefjell Ebeling, Jesse Egbert, Robert Fuchs, Sandra Goetz, Stefan Gries, Andrew Hardie, Sebastian Hoffmann, Susan Hunston, Christian Mair, Heike Pichler, Maria Stubbe, Charlotte Taylor, Gunnel Tottie and Deanna Wong. We also extend our thanks to the collaborators from CUP, Claire Dembry, Olivia Goodman, Laura Grimes, Sarah Grieves and Imogen Dickens, and the Routledge editors. Finally, we are very grateful to Sam Armstrong for assistance with formatting of the chapters.

The work on the book was supported by ESRC grants no. EP/P001559/1 and ES/K002155/1.

Vaclav Brezina
Robbie Love
Karin Aijmer
Part I

Short Introductions to Corpus-Based Sociolinguistics and the BNC2014
1 Corpus Linguistics and Sociolinguistics

Introducing the Spoken BNC2014

Vaclav Brezina, Robbie Love and Karin Aijmer

1.1 Sociolinguistics Meets Corpus Linguistics

Systematic, large-scale exploration of sociolinguistic features in everyday language use has been made possible by the availability of corpora representing informal speech, such as the demographically sampled spoken component of the British National Corpus (the Spoken BNC1994DS) and indeed the new Spoken BNC2014 (see Section 1.2). These corpora include rich metadata about social characteristics of the speakers and a large volume of data, which can be analysed using different techniques. Sociolinguistic exploration of large corpus data is, however, not without its challenges (e.g., Brezina & Meyerhoff, 2014). Language represents a dynamic system with variation occurring simultaneously at multiple levels, reflecting both conscious and unconscious choices by speakers as well as the requirements of the mode of communication, genre and a specific linguistic context (see Chapter 3 in this volume). Capturing socially meaningful variation is therefore a difficult task, requiring a good understanding of social and linguistic processes as well as familiarity with the dataset. The analysis often needs to shift between showing general patterns in the data and providing specific examples of language use to arrive at an interpretation that does justice to the complexity of the data. Bringing corpus linguistics and sociolinguistics together (cf. Baker, 2010) to investigate current spoken British English creates a unique opportunity to gain insight into everyday language use of people from different parts of the UK and different ‘corners’ of society. It is a fascinating exploration to which this volume intends to contribute.

1.2 The Spoken BNC2014: Full Dataset and Sample

For over twenty years, the British National Corpus (BNC) has been one of the most widely known corpora used as a representative sample of current British English. Focusing on the five-million-token Spoken BNC1994DS, Love, Dembry, Hardie, Brezina, and McEnery (2017) show that no other orthographically transcribed spoken corpus compiled since its release has
matched it in its size, representativeness or usefulness. However, as Love, Dembry, Hardie, Brezina, and McEnery (2017) argue, a new dataset reflecting current usage is needed to better serve the requirements of the research community than the aging Spoken BNC1994DS.

The Spoken BNC2014 is a response to this need. Publicly released in September 2017, initially via CQPweb (Hardie, 2012), the corpus is a result of collaboration between the ESRC Centre for Corpus Approaches to Social Science (CASS) at Lancaster University and Cambridge University Press (CUP). Love, Dembry, Hardie, Brezina, and McEnery (2017) describe in greater detail how the Spoken BNC2014 was designed and built within the Lancaster/Cambridge partnership; the BNC2014 user guide (Love, Hawtin, and Hardie 2017) includes information about the structure of the full 11.5-million-word corpus.

The studies in this volume are based on a five-million-token sample of the Spoken BNC2014 data, referred to as the Spoken BNC2014 Sample (Spoken BNC2014S), which contains transcripts from conversations recorded between 2012 and 2015. The Spoken BNC2014S was made available on a competitive basis to the authors of this volume, who focused on a variety of sociolinguistic applications (see Section 1.3). The Spoken BNC2014S consists of 4,784,691 tokens (including punctuation), approximately 60% of which were produced by female speakers and 40% by male speakers. A wide range of age groups are represented in the dataset, with the largest proportion (41%) in the data being produced by speakers between 19 and 29. Information is also available about the speaker’s socio-economic status and region. A detailed break-down of these categories is provided in the Appendix at the end of this chapter.

1.3 Sociolinguistic studies of the Spoken BNC2014

This volume offers four short theoretical/methodological pieces and eight empirical studies. It demonstrates a corpus-based sociolinguistic approach to the Spoken BNC2014 and provides a snapshot of sociolinguistic variation in spoken British English in the 2010s, often contrasted with the situation in the 1990s. The volume is divided into three broad sections: (i) Introductions, (ii) Discourse, Pragmatics and Interaction, and (iii) Morphosyntax.

I Introductions to Corpus-Based Sociolinguistics and the Spoken BNC2014

In addition to this introduction, the first section of this volume comprises three short contributions, which offer a reflection about the state of the art in corpus linguistics and sociolinguistics and provide context to the empirical chapters that follow. McEnery offers a compelling account of the major design decisions when building the Spoken BNC2014; this chapter lays out principles of spoken corpus design and highlights the
importance of data in corpus linguistics. Busse's contribution outlines different sociolinguistic perspectives on British English with the focus on the current debates in the field. Finally, Hardie highlights some of the main features to be found in CQPweb—the online corpus analysis system which hosts the Spoken BNC2014.

II Discourse, Pragmatics and Interaction

This section is devoted to studies dealing with language use in context and the dynamics of discourse. In Chapter 5, Culpeper and Gillings focus on a well-known stereotype about British politeness. Whilst politeness in Britain is often thought of as a monolithic phenomenon characterised by indirectness, there is an assumption in lay discourse that northerners are perceived as having very different politeness practices from southerners, practices which, broadly, are characterized by friendliness. The authors put this assumption to the test by selecting fourteen key British formulaic politeness expressions, each belonging to one of three different types of politeness (tentativeness, deference or solidarity), and then examining their frequencies in the combined north and south components of the Spoken BNC2014S and the Spoken BNC1994DS.

In Chapter 6, Aijmer draws attention to new and unusual intensifiers in present-day English which appear to be in the process of undergoing delexicalisation and grammaticalisation. The following intensifiers fit into this category of intensifiers: fucking, super, dead, real, well (good) and so(+NP), in their roles as intensifiers before adjectives. Aijmer's method involves a comparison of the intensifiers in the Spoken BNC1994DS and Spoken BNC2014S.

The aim of Axelsson’s contribution (Chapter 7) is to provide an in-depth analysis of the frequencies and formal features of tag questions (including instances with init) as well as their distribution across gender, age, dialect and socio-economic status. This study complements the evidence in her previous work, which is based on the BNC1994DS. The study thus explores diachronic change in informal discourse and its dynamics.

In Chapter 8, the final chapter in this section, Wong and Kruger examine structural categories derived from the number of words, non-words and partial forms that contribute to a backchannel. They seek to establish the factors that condition the selection of various backchannel structures in British English, using a multifactorial method. With the help of corpus annotation, they identify backchannel structures, and then use grammatical and speaker metadata associated with each utterance as predictors of backchannel choice.

III Morphosyntax

The final section in this volume is concerned with morphosyntactic features in British speech. Säily, González-Díaz and Suomela’s chapter (Chapter 9) is
a contribution which investigates the use of adjective comparison. It focuses on the productivity of two comparative strategies in English: the inflectional -er and the periphrastic more strategy. The study builds on recent research using novel methodologies that shows sociolinguistic variation in the productivity of extremely productive derivational suffixes.

Jenset, McGillivray and Rundell’s contribution (Chapter 10) investigates English verbs whose argument structure preferences include the dative alternation (Give me the money/Give the money to me). Although this is a well-researched topic, most published work draws either on introspection or on data from written sources. Using contemporary unscripted spoken data will therefore take the research into fresh territory and will bring new insights about the dative alternation in spoken English with attention being paid to sociolinguistic variation.

Caines, McCarthy and Buttery (Chapter 11) investigate zero auxiliary use in progressive aspect interrogatives in spoken British English, e.g., You talking to me? Where we going? What you been doing? The authors outline the situation of the progressive aspect in English, including the zero auxiliary, and offer two comparable empirical studies on the use of zero auxiliary in British speech, one dealing with data from the 1990s, the other with 2010s data.

Finally, Paterson (Chapter 12) explores the use of untriggered reflexives in current British English from the sociolinguistic perspective, i.e., the use of untriggered reflexives by particular demographic groups (defined by age, gender, etc.). The analysis provides a snapshot of current usage of untriggered reflexives and facilitates comparison with the existing corpus-based research of this grammatical phenomenon.

Note

1 The Spoken BNC2014 compilation project was supported by the ESRC Centre for Corpus Approaches to Social Science, ESRC grant reference ES/K002155/1.

References


Appendix: Population of the Main Speaker Demographic Categories in the Spoken BNC2014 Sample (BNC2014S)

### Gender

<table>
<thead>
<tr>
<th>Gender</th>
<th>No. Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2,872,758</td>
</tr>
<tr>
<td>Male</td>
<td>1,911,836</td>
</tr>
<tr>
<td>Unknown</td>
<td>97</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,784,691</strong></td>
</tr>
</tbody>
</table>

### Age

<table>
<thead>
<tr>
<th>Age</th>
<th>No. Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>1,281</td>
</tr>
<tr>
<td>11–18</td>
<td>191,987</td>
</tr>
<tr>
<td>19–29</td>
<td>1,961,779</td>
</tr>
<tr>
<td>30–39</td>
<td>834,379</td>
</tr>
<tr>
<td>40–49</td>
<td>463,022</td>
</tr>
<tr>
<td>50–59</td>
<td>375,368</td>
</tr>
<tr>
<td>60–69</td>
<td>625,013</td>
</tr>
<tr>
<td>70–79</td>
<td>254,263</td>
</tr>
<tr>
<td>80–89</td>
<td>45,066</td>
</tr>
<tr>
<td>90–99</td>
<td>3,812</td>
</tr>
<tr>
<td>Unknown</td>
<td>28,271</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,784,241</strong></td>
</tr>
</tbody>
</table>

### Socio-Economic Status: NS-SEC

<table>
<thead>
<tr>
<th>NS-SEC</th>
<th>Description</th>
<th>No. Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Higher managerial, administrative and professional occupations:</td>
<td>81,728</td>
</tr>
<tr>
<td>1.2</td>
<td>Large employers and higher managerial and administrative occupations</td>
<td>106,0691</td>
</tr>
<tr>
<td>2</td>
<td>Higher professional occupations</td>
<td>1,498,777</td>
</tr>
<tr>
<td>NS-SEC</td>
<td>Description</td>
<td>No. Words</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>3</td>
<td>Lower managerial, administrative and professional occupations</td>
<td>527,335</td>
</tr>
<tr>
<td>4</td>
<td>Intermediate occupations</td>
<td>95,523</td>
</tr>
<tr>
<td>5</td>
<td>Small employers and own account workers</td>
<td>93,004</td>
</tr>
<tr>
<td>6</td>
<td>Lower supervisory and technical occupations</td>
<td>78,227</td>
</tr>
<tr>
<td>7</td>
<td>Semi-routine occupations</td>
<td>40,390</td>
</tr>
<tr>
<td>8</td>
<td>Routine occupations</td>
<td>668,608</td>
</tr>
<tr>
<td></td>
<td>Uncategorised Students/unclassifiable</td>
<td>614,721</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>25,687</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4,784,691</td>
</tr>
</tbody>
</table>

**Socio-Economic Status: Social Grade**

<table>
<thead>
<tr>
<th>SG</th>
<th>Description</th>
<th>No. Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Higher managerial, administrative and professional</td>
<td>1,142,419</td>
</tr>
<tr>
<td>B</td>
<td>Intermediate managerial, administrative and professional</td>
<td>1,498,777</td>
</tr>
<tr>
<td>C1</td>
<td>Supervisory, clerical and junior managerial, administrative and professional</td>
<td>622,858</td>
</tr>
<tr>
<td>C2</td>
<td>Skilled manual workers</td>
<td>93,004</td>
</tr>
<tr>
<td>D</td>
<td>Semi-skilled and unskilled manual workers</td>
<td>118,617</td>
</tr>
<tr>
<td>E</td>
<td>State pensioners, casual and lowest grade workers, unemployed with state benefits only</td>
<td>1,283,329</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>25,687</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4,784,691</td>
</tr>
</tbody>
</table>

**Region**

<table>
<thead>
<tr>
<th>Global</th>
<th>Country</th>
<th>Supra-Region</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>Country</td>
<td>Supra-Region</td>
<td>Region</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>South (2,470,535)</td>
<td>South West</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(33,104)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>South East (not London)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(215,420)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>London (188,188)</td>
</tr>
<tr>
<td></td>
<td>Scottish (10,440)</td>
<td>Scottish (10,440)</td>
<td>Scottish (10,440)</td>
</tr>
<tr>
<td></td>
<td>Welsh (40,843)</td>
<td>Welsh (40,843)</td>
<td>Welsh (40,843)</td>
</tr>
<tr>
<td></td>
<td>Northern Irish (0)</td>
<td>Northern Irish (0)</td>
<td>Northern Irish (0)</td>
</tr>
<tr>
<td>Non-UK (74,214)</td>
<td>Irish (12,462)</td>
<td>Irish (12,462)</td>
<td>Irish (12,462)</td>
</tr>
<tr>
<td></td>
<td>Non-UK (61,752)</td>
<td>Non-UK (61,752)</td>
<td>Non-UK (61,752)</td>
</tr>
<tr>
<td>Unspecified (291,284)</td>
<td>Unspecified</td>
<td>Unspecified (655,169)</td>
<td>Unspecified (2,686,655)</td>
</tr>
</tbody>
</table>
Short Introductions to Corpus-Based Sociolinguistics and the BNC2014


sociolinguistic change (pp. 428–455). Berlin: Walter de Gruyter.
Corpus Linguistics and Sociolinguistics

The Spoken BNC2014

Current British English


Using the Spoken BNC2014 in CQPweb


Discourse, Pragmatics and Interaction


That's Well Bad


Canonical Tag Questions in Contemporary British English


Yeah, Yeah Yeah or Yeah No That’s Right


Pullin, P. (2015). The application of English as a Lingua Franca (ELF) research findings to the teaching of pragmatic competence. In K. Beeching & H. Woodfield (Eds.), Researching sociopragmatic variability: Perspectives from variational, interlanguage and contrastive pragmatics (pp. 276-296). Hampshire: Palgrave Macmillan Ltd.


Morphosyntax


Publications.


Variation in the Productivity of Adjective Comparison in Present-Day English

The Dative Alternation Revisited
Sapporo, Japan: Association for Computational Linguistics.
Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. Language Variation and Change, 24(2), 135-178. doi:10.1017/S0954394512000129

You Still Talking to Me? You Still Talking to Me?


You Can Just Give Those Documents to Myself
