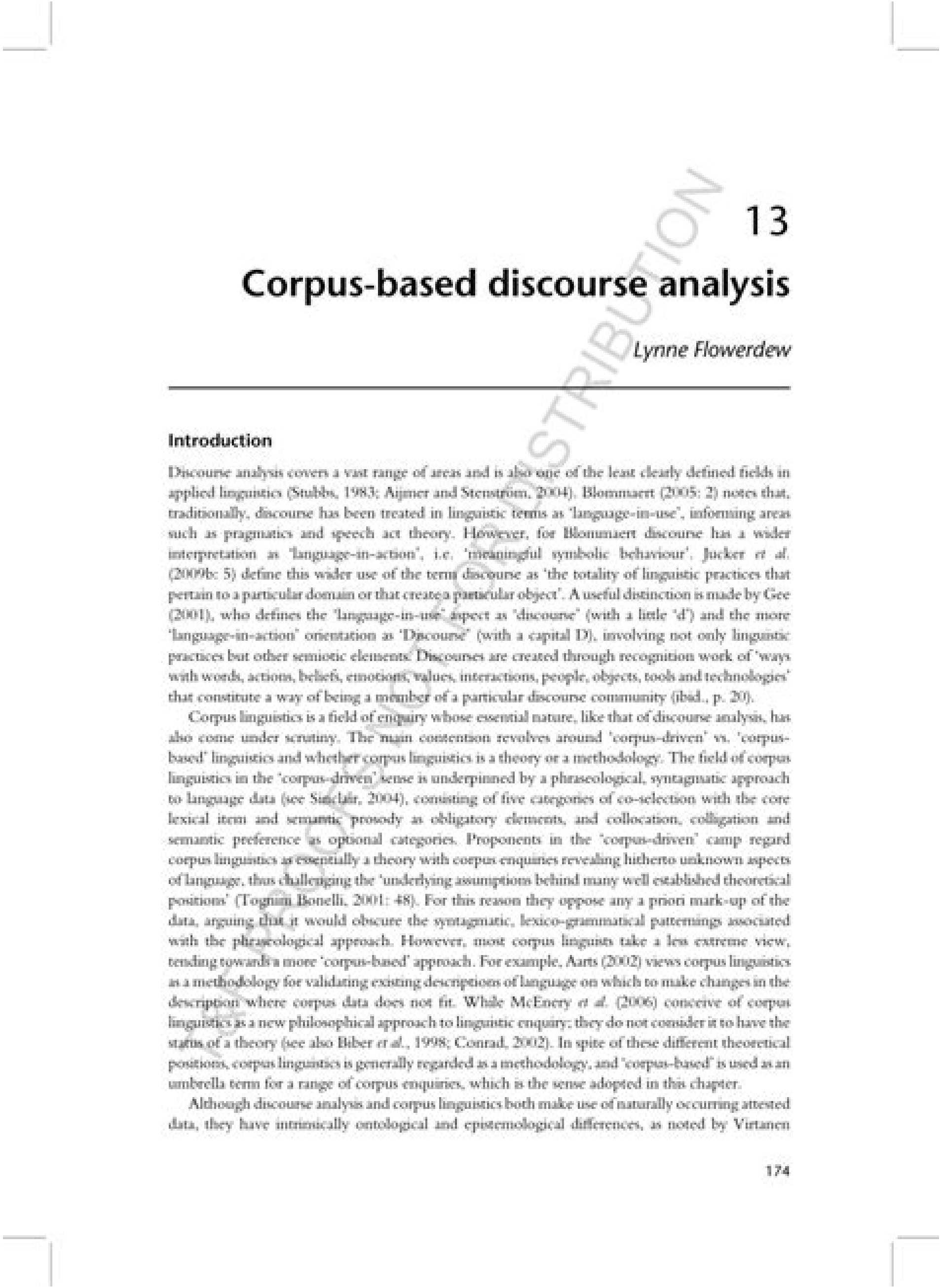
Corpus Based Discourse Analysis



13

Corpus-based discourse analysis

Lynne Flowerdew

Introduction

Discourse analysis covers a vast range of areas and is also one of the least clearly defined fields in applied linguistics (Stubbs, 1983; Aijmer and Stenström, 2004). Blommaert (2005: 2) notes that, traditionally, discourse has been treated in linguistic terms as 'language-in-use', informing areas such as pragmatics and speech act theory. However, for Blommaert discourse has a wider interpretation as 'language-in-action', i.e. 'meaningful symbolic behaviour'. Jucker *et al.* (2009b: 5) define this wider use of the term discourse as 'the totality of linguistic practices that pertain to a particular domain or that create a particular object'. A useful distinction is made by Gee (2001), who defines the 'language-in-use' aspect as 'discourse' (with a little 'd') and the more 'language-in-action' orientation as 'Discourse' (with a capital D), involving not only linguistic practices but other semiotic elements. Discourses are created through recognition work of 'ways with words, actions, beliefs, emotions, values, interactions, people, objects, tools and technologies' that constitute a way of being a member of a particular discourse community (ibid., p. 20).

Corpus linguistics is a field of enquiry whose essential nature, like that of discourse analysis, has also come under scrutiny. The main contention revolves around 'corpus-driven' vs. 'corpus-based' linguistics and whether corpus linguistics is a theory or a methodology. The field of corpus linguistics in the 'corpus-driven' sense is underpinned by a phraseological, syntagmatic approach to language data (see Sinclair, 2004), consisting of five categories of co-selection with the core lexical item and semantic prosody as obligatory elements, and collocation, colligation and semantic preference as optional categories. Proponents in the 'corpus-driven' camp regard corpus linguistics as essentially a theory with corpus enquiries revealing hitherto unknown aspects of language, thus challenging the 'underlying assumptions behind many well established theoretical positions' (Tognini Bonelli, 2001: 48). For this reason they oppose any a priori mark-up of the data, arguing that it would obscure the syntagmatic, lexico-grammatical patternings associated with the phraseological approach. However, most corpus linguists take a less extreme view, tending towards a more 'corpus-based' approach. For example, Aarts (2002) views corpus linguistics as a methodology for validating existing descriptions of language on which to make changes in the description where corpus data does not fit. While McEnery *et al.* (2006) conceive of corpus linguistics as a new philosophical approach to linguistic enquiry; they do not consider it to have the status of a theory (see also Biber *et al.*, 1998; Conrad, 2002). In spite of these different theoretical positions, corpus linguistics is generally regarded as a methodology, and 'corpus-based' is used as an umbrella term for a range of corpus enquiries, which is the sense adopted in this chapter.

Although discourse analysis and corpus linguistics both make use of naturally occurring attested data, they have intrinsically ontological and epistemological differences, as noted by Virtanen

Corpus based discourse analysis is an innovative approach that merges the principles of discourse analysis with the methodologies of corpus linguistics. This interdisciplinary field provides researchers with a robust framework for examining language use in various contexts. By utilizing large collections of texts, or corpora, researchers can uncover patterns, trends, and nuances that may not be apparent through traditional qualitative methods alone. This article delves into the key concepts, methodologies, applications, and challenges associated with corpus based discourse analysis.

Understanding Corpus Based Discourse Analysis

At its core, corpus based discourse analysis involves the systematic examination of language in use through the lens of a corpus. The concept can be broken down into two main components:

1. Corpus Linguistics

Corpus linguistics is the study of language as expressed in corpora (bodies of text) and is characterized by the following features:

- Data-Driven: Corpus linguistics relies on actual language use rather than hypothetical examples.
- Quantitative and Qualitative: It employs both statistical methods to analyze language patterns and qualitative approaches to interpret those patterns.
- Large Datasets: It utilizes extensive datasets, which allow for more comprehensive analyses than would be possible with smaller samples.

2. Discourse Analysis

Discourse analysis focuses on the study of language in context, examining how language constructs meaning in social interactions. Key aspects include:

- Contextualization: Understanding how context influences the interpretation of text and talk.
- Power Dynamics: Analyzing how language reflects and shapes social power relations.
- Intertextuality: Exploring the connections between texts and how they influence one another.

By combining these two approaches, corpus based discourse analysis offers a powerful methodology for investigating language phenomena.

Methodologies in Corpus Based Discourse Analysis

The methodologies employed in corpus based discourse analysis can vary widely depending on the research questions and the nature of the corpus being analyzed. However, several common steps and techniques are frequently used:

1. Corpus Compilation

The first step in corpus based discourse analysis is selecting and compiling a corpus that is relevant to the research question. This involves:

- Defining the Scope: Determining the thematic focus, time frame, and type of texts to include (e.g., spoken, written, formal, informal).
- Collecting Data: Using various sources such as books, articles, transcripts, social media, or interviews.
- Ensuring Representativeness: Creating a corpus that accurately represents the language variety being studied.

2. Corpus Annotation

Once the corpus is compiled, it may be necessary to annotate the data to facilitate analysis. Annotation can include:

- Linguistic Tags: Marking parts of speech, syntactic structures, or discourse markers.
- Pragmatic Annotations: Identifying speech acts, politeness strategies, or interjections.
- Contextual Information: Adding metadata about the texts, such as speaker demographics or situational context.

3. Data Analysis Techniques

Data analysis in corpus based discourse analysis can employ various techniques, including:

- Frequency Analysis: Counting the occurrence of specific words, phrases, or structures.
- Concordance Analysis: Examining the context in which a word or phrase appears (e.g., using a concordance tool).
- Collocation Analysis: Identifying words that frequently occur together and exploring their semantic relationships.
- Keyword Analysis: Finding words that are statistically significant in the corpus compared to a reference corpus.

Applications of Corpus Based Discourse Analysis

Corpus based discourse analysis has a wide range of applications across different fields, including:

1. Sociolinguistics

In sociolinguistics, corpus based discourse analysis can reveal how language varies across different social groups. Researchers can examine:

- Language Variation: How dialects and sociolects manifest in written and spoken discourse.
- Identity Construction: How individuals use language to construct and negotiate their identities in various contexts.

2. Media Studies

In media studies, researchers can analyze how language is used in media texts to shape public perception and discourse. This includes:

- Framing Analysis: Investigating how language frames issues or events in specific ways.
- Representation Studies: Analyzing how different groups or themes are represented in media discourse.

3. Education

In the field of education, corpus based discourse analysis can be applied to:

- Curriculum Development: Informing the design of teaching materials based on authentic language use.
- Language Acquisition: Understanding how learners interact with language in real-world contexts.

Challenges in Corpus Based Discourse Analysis

While corpus based discourse analysis offers significant advantages, it is not without its challenges. Some of the key challenges include:

1. Data Quality and Representativeness

- Bias in Data Selection: The corpus may not represent all variations of language, leading to biased conclusions.
- Outdated Data: Language evolves, and using outdated corpora may not reflect current usage patterns.

2. Complexity of Analysis

- Interpretative Challenges: While quantitative data can reveal patterns, interpreting those patterns within their social context requires nuanced understanding and careful analysis.
- Tools and Skills: Researchers need proficiency in both linguistic analysis and computational tools, which may not be present in all research teams.

3. Ethical Considerations

- Privacy Concerns: Using data from interviews or social media may raise ethical issues regarding consent and privacy.
- Representation Issues: Ensuring that the voices of marginalized groups are represented in the corpus analysis.

Conclusion

In conclusion, corpus based discourse analysis represents a dynamic intersection of quantitative and qualitative methodologies, allowing researchers to delve into the complexities of language use across various contexts. By leveraging large datasets, researchers can uncover patterns and insights that are essential for understanding the intricate relationship between language, society, and culture. Despite the challenges associated with this approach, its applications in fields such as sociolinguistics, media studies, and education demonstrate its value as a powerful tool for linguistic and discourse analysis. As technology continues to advance, the potential for corpus based discourse analysis will only grow, offering deeper insights into the ever-evolving landscape of language.

Frequently Asked Questions

What is corpus-based discourse analysis?

Corpus-based discourse analysis is an approach that utilizes large collections of written or spoken texts (corpora) to study language use in context, focusing on how discourse is structured and how meaning is constructed in communication.

How does corpus-based discourse analysis differ from traditional discourse analysis?

Unlike traditional discourse analysis, which often relies on qualitative methods and small samples, corpus-based discourse analysis employs quantitative methods and large data sets to identify patterns and trends in language use across different contexts.

What types of corpora are commonly used in corpus-based discourse analysis?

Common types of corpora include written texts (like newspapers, books, and academic articles), spoken texts (like conversations, interviews, and speeches), and specialized corpora focused on specific genres or communities.

What tools are commonly used in corpus-based discourse analysis?

Tools like AntConc, NVivo, and Sketch Engine are commonly used for analyzing corpora, allowing researchers to perform word frequency analysis, concordance searches, and collocation studies.

What are some key benefits of using corpus-based discourse analysis?

Key benefits include the ability to analyze large amounts of data for more representative findings, uncovering patterns that may not be visible in smaller samples, and enhancing the objectivity of linguistic analysis through statistical methods.

Can corpus-based discourse analysis be applied to social media texts?

Yes, corpus-based discourse analysis can be effectively applied to social media texts, allowing researchers to study language use, discourse practices, and the construction of identity and community in online interactions.

What is the role of context in corpus-based discourse analysis?

Context plays a crucial role in corpus-based discourse analysis as it helps researchers understand how language choices reflect social, cultural, and situational factors, influencing meaning and interpretation.

How can corpus-based discourse analysis contribute to language teaching and learning?

It can inform language teaching by providing insights into authentic language use, helping educators design materials that reflect real-world communication patterns and improve learners' understanding of discourse conventions.

What challenges are associated with corpus-based

discourse analysis?

Challenges include the need for robust corpus design, the potential for misinterpretation of quantitative data without qualitative insights, and the complexity of analyzing multimodal texts that combine various forms of communication.

What future trends are emerging in corpus-based discourse analysis?

Emerging trends include the integration of machine learning and artificial intelligence for more sophisticated analysis, the exploration of multimodal discourse in digital environments, and a greater emphasis on ethical considerations in data collection and analysis.

Find other PDF article:
https://soc.up.edu.ph/13-note/pdf?trackid=orb00-6207&title=code-signal-general-coding-assessment.pdf

Corpus Based Discourse Analysis

□□□□Neural Corpus Indexer□Camera Ready□□ ...
□□□□Neural Corpus Indexer□Camera Ready□□□□□□□□□□□ □□□□□NeurIPS2022 Outstanding Paper A ...

□□□□□□□ | □□□□□□□□
Jan 4, 2016 · □□□□□□COTE corpus (COTE Corpus of Translational English)□□ xujiajin 2022-09-26 □□ ...

Crown/CLOB□□□:2009□□□□□□□□ - □□□□□□□□
hyy 2013-02-02 #4 □□: Crown/CLOB□□□:2009□□□□□□□□ Thanks a lot, Doctor Xu. You are always so ...

Some free online Chinese corpora | □□□□□□□□
Jun 16, 2005 · This paper presents an introduction to concordancers, and to the concordancing of Chinese e-texts ...

□□□□□□□□□ | □□□□□□□□
Jun 7, 2013 · Spanish Treebank Corpus 1.500 sentences from newspapers, syntactically annotated C-ORAL-ROM ...

□□□□*Neural Corpus Indexer*□*Camera Ready*□□□□□□□□ ...
□□□□Neural Corpus Indexer□Camera Ready□□□□□□□□□□□ □□□□□NeurIPS2022 Outstanding Paper A Neural Corpus Indexer for Document Ret...

□□□□□□□ | □□□□□□□□
Jan 4, 2016 · □□□□□□COTE corpus (COTE Corpus of Translational English)□□ xujiajin 2022-09-26 □□ 3

□□ 5K 2024-12-01

Crown/CLOB□□□:2009□□□□□□□□ - □□□□□□□□
hyy 2013-02-02 #4 □□: Crown/CLOB□□□:2009□□□□□□□□ Thanks a lot, Doctor Xu. You are always so generous and supportive for all corpus lovers. Hope some days we could ...

Some free online Chinese corpora | □□□□□□□□
Jun 16, 2005 · This paper presents an introduction to concordancers, and to the concordancing of Chinese e-texts in particular. Demonstrations are given of searches using spaced and non ...

□□□□□□□□□ | □□□□□□□□
Jun 7, 2013 · Spanish Treebank Corpus 1.500 sentences from newspapers, syntactically annotated C-ORAL-ROM Multilingual Spoken Corpus (Spanish, French, Portuguese and ...

□□□□□□□□□□□□ | □□□□□□□□
Apr 14, 2016 · College Learners Spoken English Corpus (COLSEC)

□□□□□□□□□□□□*data set* □ *dataset* □□□□□□□□ *...*
Nothing from before then. Interestingly, the British National Corpus has 51 incidences, dating from the 1980s to the mid 1990s. dataset□□□□□□□□□□ □□□□□□□□□□□□□□ ...

□□□□□□□□□□□□□□□ - □□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□K□□□ OPUS - an open source parallel corpus□□□□□□□□□□□□□□□□□□□□□□□□□□□□TED□□□□□□ ...

The BNC is now freely downloadable | □□□□□□□□
In 2014 there will be some changes in the way that the British National Corpus (BNC) is distributed. It is now possible to download the British National...

□□□□□ | □□□□□□□□
Mar 19, 2025 · Chinese Corpora□□□wordsmith4□□□□□□□□□□□□□□□□□□□

Explore the essentials of corpus-based discourse analysis and its powerful applications in language research. Discover how this method can enhance your findings!

Back to Home