



CZECH NATIONAL  
CORPUS



Introduction to Text Corpora and Their Applications

# Corpora in discourse studies (MD-CADS)

Lucie Chlumská, Ph.D.

[lucie.chlumska@korporus.cz](mailto:lucie.chlumska@korporus.cz)





# OUTLINE:

## 1. LECTURE

- discourse analysis and corpus linguistics: distinction
- methods used in CADS
- case studies in CADS
  - popular topics (gender, race, immigration, politics...)

## 2. SEMINAR

- reading (Partington & Marchi): *Corpora in discourse analysis*
- forced priming – what is it and how can it be studied?
- can CADS be objective (as opposed to CDA)?





# LECTURE





# Discourse analysis



# What is discourse?

structural definition:

„language above the sentence or above the clause“

functional definition:

„language that is doing some job in some context“

- an instrument put to work:
  - to inform
  - to influence the ideas, opinions, and behaviour of other communication participants



# Discourse analysis

- usually referring to **critical discourse analysis** (CDA)
  - concerned with the studies of ideology, power, and culture
  - relies heavily on real language data, but also on the analyst's knowledge and experience (incl. prejudices...)
  - often politically driven purpose
- example from the **Czech context**:
  - CDA of statements concerning the Šumava National Park
    - advocates of logging and further interventions
    - advocates of preserving the current state or extension of the protected areas



# CDA v. CL

- cultural divide between CL and (C)DA (Leech 2000)

## (C)DA

- emphasizes the integrity of text
- primarily qualitative
- focuses on the contents expressed by language
- the collector, transcriber and analyst are often the same person
- data rarely widely available

## CL

- often uses representative samples
- essentially quantitative
- focuses primarily on the language *per se*
- rarely the case in corpus linguistics
- data typically made widely available





# CADS v. CDA

- **corpus-assisted discourse studies**  
= the set of studies into the form and/or function of language *as communicative discourse* which incorporate the use of computerised corpora in their analyses
- the term CADS first coined by Partington (2004), but much CADS research was being conducted before that
- **CADS** is *not* tied to any particular school of discourse analysis, certainly not to **critical discourse analysis (CDA)**
  - unlike CDA, it has no overreaching political agenda and has very different attitudes to and traditions of how language data should be treated



# Aim of CADS

- the main objective:
  - to uncover, in the discourse type under study, **non-obvious meaning**, i.e. to reveal something that might not be visible to the naked-eye perusal





# Methods used in CADs



# Combination of methods

- not a single favourite method, rather their combination
- traditional DA method:
  - close reading
    - thoughtful, critical analysis of a text that focuses on significant details or patterns in order to develop a deep, precise understanding of the text's form, craft, meanings
- corpora:
  - concordances
  - filtering
  - wordlists
  - keywords



# Corpora used

- ad hoc compilation of **specialized corpora**
  - often corpora are not available of the discourse type in question
  - typically, corpora of various transcriptions are used (spoken language regarded as a discourse type)
  - single discourse v. heterogeneric corpora (BNC)
- CADS are essentially **comparative**
  - it is only possible to both uncover and evaluate the particular features of a discourse type by comparing it with others





# Case studies in CADs



# Popular topics

- **political discourse:**
  - **Teubert (2000)**
    - language of euroscepticism in Britain on the basis of texts from anti-EU websites (based on collocation and phraseology)
    - implicit statement: only Britain is a true democracy with a truly accountable government
  - **Fairclough (2000)**
    - comparative analysis of keywords in a corpus of Tony Blair's speeches, showing the change in attitude and Labour stance (Old v. New Labour)



# Popular topics

- minorities:
  - Krishnamurthy (1996)
    - *ethnic, racial and tribal*
    - *racial discrimination (Done-to) v. ethnic violence and tribal killing*  
(Doer)
  - RASIM by Lancaster university (Baker, McEnery)
    - refugees, asylum seekers, immigrants in the UK press and UN documents
    - analysis of collocates, cooperation of corpus linguists and CDA (?)





# Popular topics

- gender:
  - Pearce (2008)
    - differences between the lemmas *man* and *woman*, based on their collocational profile > gender stereotypes
  - Baker (2006, 2008)
    - compares *spinster* and *bachelor* in the BNC
    - *eligible bachelors* v. *frustrated spinsters*
  - Taylor (2013)
    - boy/s and girl/s in the British press from 1993-2010



Thank you for your attention!

Questions?





# SEMINAR



# Reading

common reading:

Partington, A. & Marchi, A. (2015). Using corpora in discourse analysis. In Biber, D. & R. Reppen (Eds.), *Handbook of Corpus Linguistics*. (pp. 216-234). Cambridge: Cambridge University Press.



# Discussion

- What might be an interesting topic for CADS in your home country/culture?
- What is forced priming?
- How do prefabricated phrases in the White House briefings corpus differ from general spoken discourse?
- How can the Podium influence the public picture of the White House? How can we study that?
- How can we identify what is absent in a corpus/discourse?

