CZECH NATIONAL
CORPUS

CZECH NATIONAL CORPUS

Introduction to Text Corpora and Their Applications

Corpora in stylistics and literary studies

Lucie Chlumská, Ph.D.

lucie.chlumska@korpus.cz

OUTLINE:

1. LECTURE

- corpus stylistics
- style and literary language
- methods in corpus stylistics
- case studies in corpus stylistics

2. SEMINAR

- reading (Jonathan Culpeper): *Keyness in Romeo and Juliet*
- keywords: what can they reveal about the text and style?

LECTURE

Corpus stylistics

What is corpus stylistics?

Leech (2008): „the study of style is essentially the study of *variation* in the use of language“

Corpus stylistics = the study of literary texts that employs corpus-linguistics methods to support the analysis of textual meanings and the interpretation of texts

corpus-stylistic research can focus on individual texts and even text extracts as the places where the aesthetic effects of language are best analyzed

Basis of corpus-stylistic research

- intrinsic explanatory purpose of the linguistic analysis
- Leech (2008: 54): descriptive v. explanatory stylistics
  - descriptive: purpose is to describe the style
  - explanatory: to use stylistics to explain something

- explanatory goal may be extrinsic or intrinsic
  - extrinsic: to identify the author of a text or the chronological relationship between the texts
  - intrinsic: to explain the meaning of the text

- the intrinsic explanatory purpose of corpus stylistics > close to literary stylistics (linguistic analysis + literary criticism)
- also an extrinsic dimension: compares texts and assesses specific linguistic features in relation to wider linguistic patterns

Literary style

- literary style as an object of study is difficult to define > rather „meanings in literary texts“
- style is closely allied to registers/genres and dialects/language varieties
- stylistic shifts in usage may be observed with reference to features associated with either particular situations of use or particular groups of speakers

- style mostly associated with literary texts > literary language
    - Burrows (1987): disctinction between literary and non-literary language is not a clear-cut one
    - not a different set of features, but rather a continuum
    - literary texts often defined extralinguistically (publishers, libraries)

Literary style

- literary texts in corpora tend to be analyzed from a register view
  - Biber et al. (1999): four registers (fiction, news, conversation and academic prose)
  - register perspective studies frequent and pervasive words and grammatical structures and interprets them in respect of situational characteristics of the variety
  - focus on frequent features
  - patterns resulting from the real-world situational characteristics of the registers are functional

- a genre approach, in contrast, requires complete texts
  - features associated with styles are not functional, but rather associated with aesthetic preferences
  - focus on specific features, not just the frequent ones

Methods in corpus stylistics

Corpus-stylistic methods

- researchers usually employ standard functionalities offered by concordancers:
  - retrieving frequencies
  - analyzing concordances
  - generating collocations
  - extracting keywords

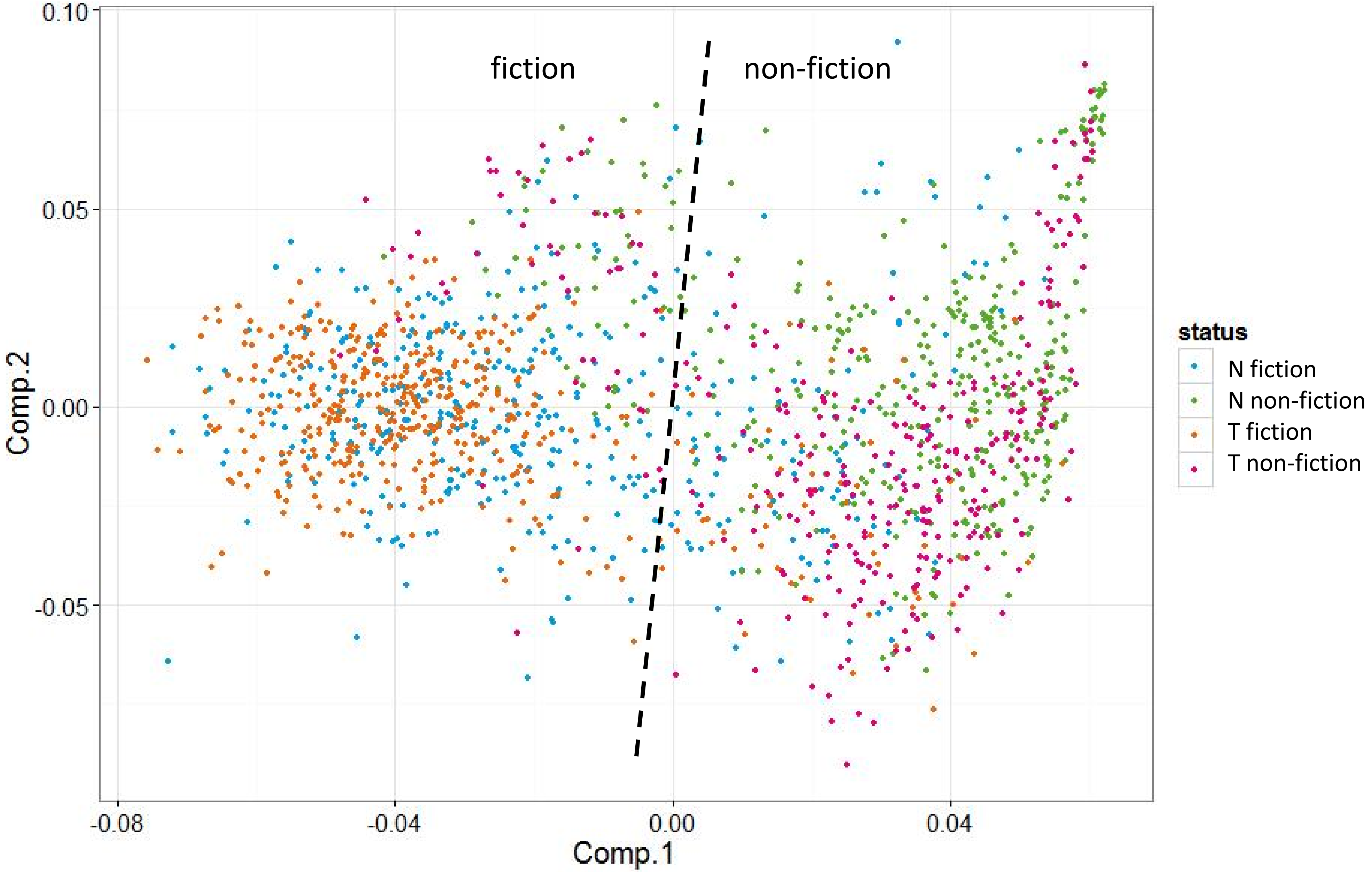
- one of the often used methods is PCA (principal component analysis):
  - statistical procedure reduces sets of variables to a smaller number of composite variables to account for most of the relationship between initial variables
  - can depict the variability of the data and show which texts or sets of data differ the most or least

Corpus-stylistic methods

- often, sophisticated statistical methods are used:
  - PCA (principal component analysis)
  - MDA (Biber’s multidimensional analysis)
  - cluster analysis...

- combination of stylistics and statistics gave birth to a new interdisciplinary area called stylometry, computational staylistics or statistical stylistics
  - often used to study the authorial style of individual authors
  - also in forensic lingistics (to attribute authorship of anonymous texts)

PCA: genre differences



Case studies

PCA-based studies

- Craig (2008): used PCA to study relationships between Shakespeare’s characters
  - focuses on 50 characters that speak more than 3,000 words and uses the fifty most frequent words
  - interprets the PCA results as „a sociolinguistics of character“
  - shows contrasts and similarities between characters that seem to reflect the social purposes of the character’s speech
    - female characters use *I* and *me* more than *we* and *our*, reflecting individuality

- Tabata (1995): used PCA to compare first-person and third-person narratives in Dickens (100 most common words, 9 novels)
  - 1st person: verbal structures, *and*, *but*, intensifiers, negatives
  - 3rd person: nominal structures with *the*, *which*, *who*, actions

Frequent words

- Burrow (1987): important contribution to showing how empirical evidence can add objectivity to the study of textual features
  - shows that frequent words play an important role in the creation of meaning in a novel
  - top frequent, mostly grammatical, words usually do not receive particular attention
  - Burrow argues that the top frequent words reflect unobtrusive habits of expression and can differentiate idiolects of characters

    - Jane Austen’s language: her dialogues with other novelists
    - focus on methodology, rather than findings (small range of examples)
    - another important observation: literary language cannot be so easily distinguished from non-literary language

Frequent words

- Stubbs (2005): in his study, he tries to „illustrate the literary value of simple quantitative text and corpus data“
  - seeks the links between corpus findings and literary interpretation
  - study of Conrad’s *Heart of Darkness*
    - first, brief outline of some observations put forward by literary critics (theme of unreliable knowledge)
    - then, he relates them to linguistics features identified with corpus methods
    - critics focused only on content words (*fog*, *indistinct*) as expressions of vagueness and disregarded grammatical structures such as *something*, *somewhere*, *kind of*, *sort of*.
  - Stubbs argues that corpus-based techniques can add systematicity and detail to textual analysis and literary interpretation

Semantic prosody

- Louw (1993): illustrates the potential of the concept of „semantic prosody“ in its application to stylistics
- practise of „matching texts against corpora“ as a useful method of corpus stylistics
  - semantic prosodies result from habitual collocates colouring the meanings of words they occur with
  - semantic prosody of melancholia: *days are* (*gone*, *over*, *past*...)
  - theme of Larkin’s poem „Days“ > the line „Days are where we live“ triggers associations of melancholia that point forward to the theme of death developed in the poem
  - semantic prosody as a „consistent aura of meaning with which a form is imbued by its collocates“ provides a background to these expectations

Thank you for your attention!

Questions?

CZECH NATIONAL CORPUS

SEMINAR

Reading

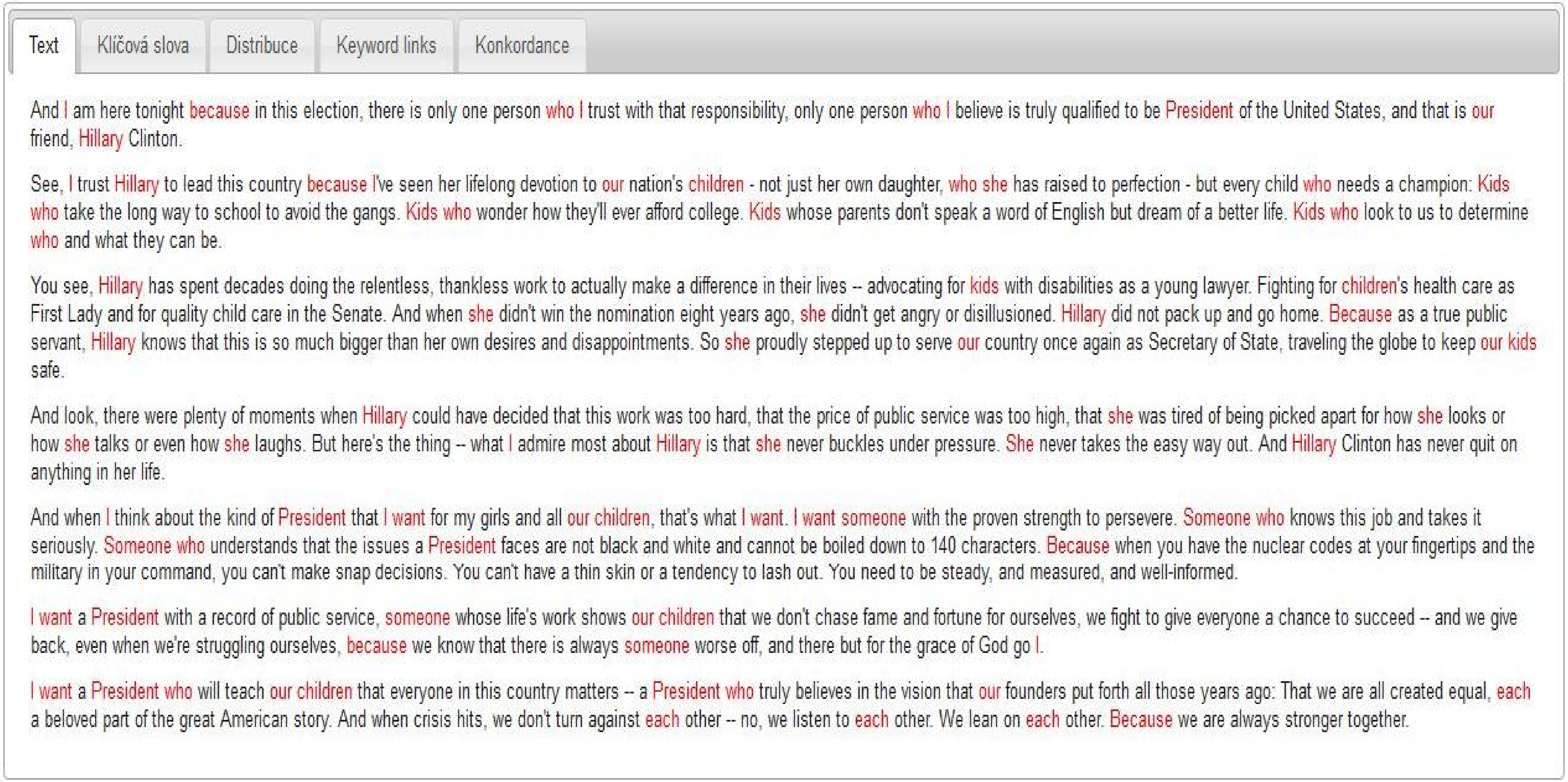
common reading:

Culpeper, J. (2009). Keyness. Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's *Romeo and Juliet*. *International Journal of Corpus Linguistics*. 14:1, 29–59.

Discussion

- What is a keyword?
- How can semantic tagging be useful in keyword analysis?
- What is a style marker (according to Nils Erik Enkvist)?
- Whow can keywords be extracted from a corpus?
- Why does the choice of a reference corpus matter?
- What is a „cut-off point“?
- Are there more types of keywords?
- What is „aboutness“?

Who is speaking...?



And I am here tonight because in this election, there is only one person who I trust with that responsibility, only one person who I believe is truly qualified to be President of the United States, and that is our friend, Hillary Clinton.

See, I trust Hillary to lead this country because I've seen her lifelong devotion to our nation's children - not just her own daughter, who she has raised to perfection - but every child who needs a champion: Kids who take the long way to school to avoid the gangs. Kids who wonder how they'll ever afford college. Kids whose parents don't speak a word of English but dream of a better life. Kids who look to us to determine who and what they can be.

You see, Hillary has spent decades doing the relentless, thankless work to actually make a difference in their lives -- advocating for kids with disabilities as a young lawyer. Fighting for children's health care as First Lady and for quality child care in the Senate. And when she didn't win the nomination eight years ago, she didn't get angry or disillusioned. Hillary did not pack up and go home. Because as a true public servant, Hillary knows that this is so much bigger than her own desires and disappointments. So she proudly stepped up to serve our country once again as Secretary of State, traveling the globe to keep our kids safe.

And look, there were plenty of moments when Hillary could have decided that this work was too hard, that the price of public service was too high, that she was tired of being picked apart for how she looks or how she talks or even how she laughs. But here's the thing -- what I admire most about Hillary is that she never buckles under pressure. She never takes the easy way out. And Hillary Clinton has never quit on anything in her life.

And when I think about the kind of President that I want for my girls and all our children, that's what I want. I want someone with the proven strength to persevere. Someone who knows this job and takes it seriously. Someone who understands that the issues a President faces are not black and white and cannot be boiled down to 140 characters. Because when you have the nuclear codes at your fingertips and the military in your command, you can't make snap decisions. You can't have a thin skin or a tendency to lash out. You need to be steady, and measured, and well-informed.

I want a President with a record of public service, someone whose life's work shows our children that we don't chase fame and fortune for ourselves, we fight to give everyone a chance to succeed -- and we give back, even when we're struggling ourselves, because we know that there is always someone worse off, and there but for the grace of God go I.

I want a President who will teach our children that everyone in this country matters -- a President who truly believes in the vision that our founders put forth all those years ago: That we are all created equal, each a beloved part of the great American story. And when crisis hits, we don't turn against each other -- no, we listen to each other. We lean on each other. Because we are always stronger together.