

Corpus-based critical discourse analysis in translation studies

Kim
ch.4



Introduction

- CDA → language as a powerful social tool
- Wodak & Meyer → language functions as a sophisticated mechanism
- CDA × objectivity,
- CDA = qualitative, close reading
- CDA → complemented by other methods
- CAT = solution !!
- Software → uncover co-occurrence of specific words with particular social issues

Incorporating corpora into CDA =

- reducing subjectivity of text selection
- combining rich qualitative interpretation with statistically supported evidence.



- Corpus linguistics → ensures transparency in sampling + adds empirical weight to CDA interpretations.
- Such a combined approach can reveal linguistic tendencies that human intuition alone may overlook.
- Translation studies, which already embraces CAT , has naturally adopted corpus-based CDA.
- Applied to translation and interpreting
- Integrated method → deeper insights into ideological patterns and linguistic choices across languages.

CBCDA = valuable analytical framework within translation research.

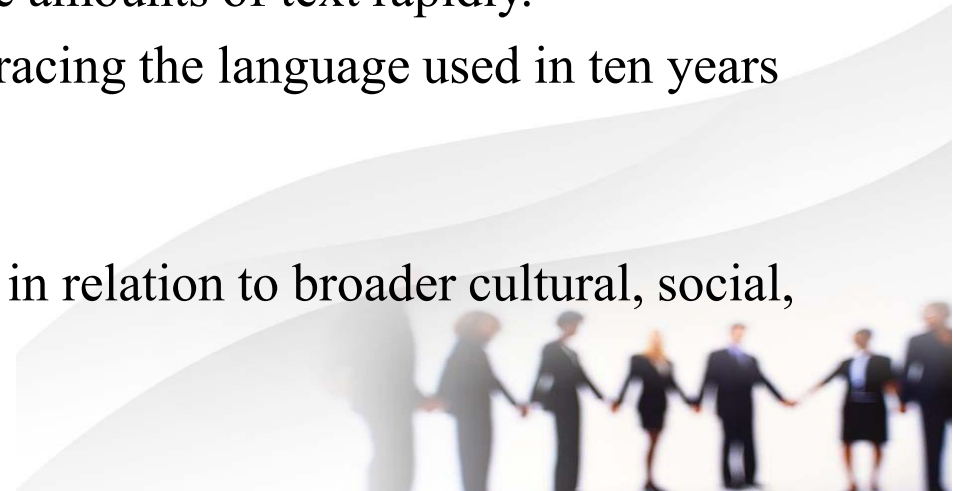


Corpus- based CDA: the benefits

- Gap between quantitative and qualitative approaches
- Corpus-based methods + CDA = researchers attempt to achieve balance
- Corpus-based approaches/quantitative→ collocational patterns + recurrent lexical behaviours
- CDA→ close reading + interpretation + examination of ideological and discursive features.
- When integrated = researchers can move beyond intuition and develop well-supported generalisations

- Key advantage of corpus-based methods = ability to process large amounts of text rapidly.
- allowing researchers to conduct extensive longitudinal studies (tracing the language used in ten years of news reporting on a particular political topic).

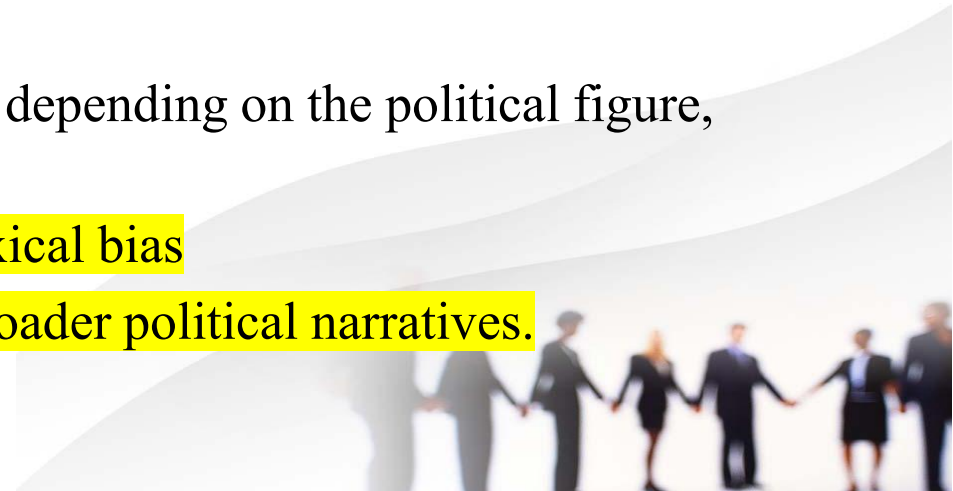
- Once patterns are identified quantitatively, they can be examined in relation to broader cultural, social, and economic contexts.



- Integration → facilitates comparison of journalistic practices across media outlets.
- Corpus analysis of references to Kim Jong Il from 1998 to 2010 across CNN, Newsweek, and The New York Times reveals striking differences in lexical choices.
- Kim Jong Il is rarely labelled “President”: out of 7,446 concordance lines, only 120 instances (2%) use this title, with Newsweek using it the least.
- By contrast, the late South Korean president Kim Dae Jung is overwhelmingly referred to as “President”, appearing with this title 880 times (69% of all occurrences).
- Objective statistics → media institutions select terms differently depending on the political figure, shaping public perceptions.

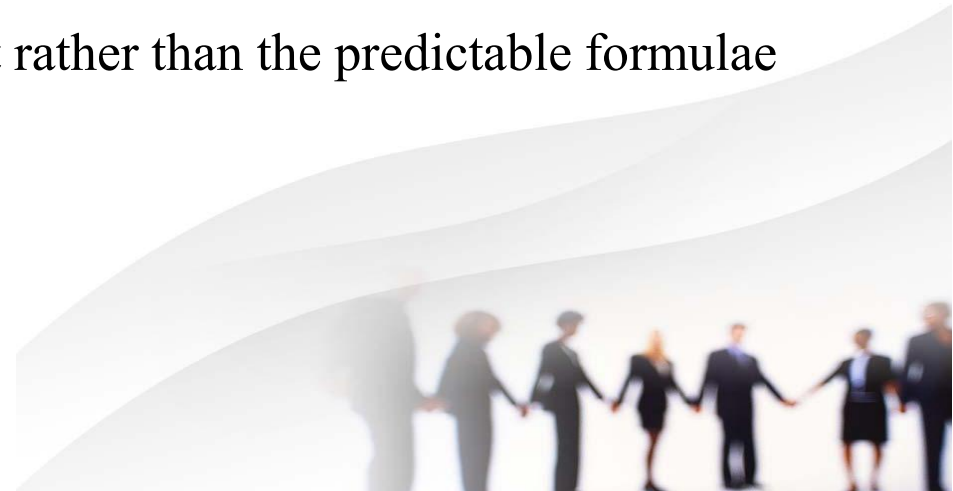
Quantitative → provide evidence of lexical bias

Qualitative → explains how these choices align with broader political narratives.



Analytical methodology for corpus-based CDA research: Analysing frequency lists

- Key starting point in CBDA = construction & examination of a frequency list
- Such lists → valuable point of entry into the data ≠ they cannot stand alone;
- In English-based DS → 1st step = removing high-frequency function words
- Excluded → stop lists, which corpus tools such as WordSmith can apply automatically.
- Genre also shapes the refinement process.
- News texts → contain dates, section headings and repeated structural elements that may artificially raise the frequency of particular items.
- must be filtered out → list reflects the meanings of the main text rather than the predictable formulae of news production.



- Once refined, frequency list → powerful resource for examining content words
- It → complementary evidence in translation studies, helping confirm interpretations made through qualitative analysis

Example:

- Zanettin, Saldanha, and Harding (2015) generated a frequency list from the Translation Studies Abstracts (TSA) online database and compared it to a larger reference corpus.
- After filtering out function words, the refined list revealed that translation, language, and English ranked as the top three content words, empirically confirming the dominance of English and text-based approaches within the discipline.



- imp→ frequency lists = preliminary diagnostic tools, not final interpretive instruments.
- Frequency alone ≠ guarantee representativeness.
- A word may appear often without forming meaningful associations with key terms in the corpus.
- Frequency lists + precise statistical measures (collocation analysis) = uncover the propensity of words to appear together.

A frequency list is best viewed as an initial observation point.



Analysing a collocate list

- Central analytical tool in CBDA.
- Scott (2022) → collocates as words that systematically appear near a search term within a corpus.
- Co-occurrence more frequently = collocation
- Meaning = greater than the sum of the individual words.
- Collocation therefore reveals patterned, socially loaded linguistic behaviour that can provide insight into discourse, ideology, and representation.



- Milani's (2013) analysis of online profiles reveal ideologically charged distinctions in how identities are linguistically constructed within a community.
- For critical discourse analysts, collocational profiles are particularly valuable because they can expose how groups, events, or social phenomena are routinely framed.
- Baker, Gabrielatos, and McEnery's (2013) study of the collocates of Muslim in British newspapers (1998–2009), using Sketch Engine?
- Community/world collocates → contributed to the construction of “Muslims” as a unified\threatening
- Highlighting implicit ideological patterns that would be far less evident through impressionistic reading.



- Examining collocate profiles of a node word in ST and its translation in TT:
- Researchers → identify shifts in discourse, ideology, or representation introduced through translation.
- Differences in collocational behaviour → reveal translator intervention, cultural filtering, or ideological reframing.
- Collocate lists → a systematic method for understanding how discourse travels, or transforms across languages.



Fairclough's New Labour Study

- Compared New Labour vs. Old Labour corpora.
- Found repeated key terms: we, new, Britain, reform.
- Frequent pairing of new + political ideas (new Labour, new politics) → creates a modern, transformative image.
- Differences between modernisation (UK-focused) and reform (EU/economic context).
- Shows how collocation reveals ideological framing.



What Collocates Tell Us

- Collocates show recurrent lexical associations.
- Analysts filter out function words to keep meaningful content words.
- Collocates can be grouped semantically to see patterns (e.g., countries, objects).
- Collocational analysis exposes discourses, ideology, and framing.

Use in Translation Studies

- Compare ST collocates vs. TT collocates.
- Reveals shifts, ideological filtering, or translator intervention.
- Shows how discourse changes across languages.

Collocational analysis provides empirical grounding for CDA.

