

Course Title: Issues in Translation Studies  
Course Tutor: Prof. Ahmed Qadouri (Ph. D.)  
PhD Programme in Translation 2025-2026  
PPT by: Farah A. Abo Al-Timen  
PhD Candidate

Corpus-based critical discourse analysis in translation studies

# Use of Chinese Characters in Korean Media & Its Impact on Corpus Analysis

- ▶ Korean media often combines Hangul with Chinese characters (Hanja) to avoid ambiguity and save space.
- ▶ Many Korean words share the same pronunciation, so adding Hanja (e.g., 李 in 이 대통령) clarifies meaning.
- ▶ Hanja is more compact (e.g., 美 for “the US,” 訪韓/訪美 for “visit Korea/the US”).
- ▶ In corpora, Hangul and Hanja forms are counted separately, causing inaccurate ST–TT comparisons.
- ▶ Example: “visit” may appear 100 times in ST, but only 40 as 방문 in TT because the rest occur as 訪韓/訪美.

# Handling Korean Morphology in Corpus Collocation Analysis

- ▶ Corpus tools treat word forms mechanically, so extra preprocessing is needed for languages like Korean.
- ▶ Korean nouns take many particles and suffixes (e.g., -은/는, -이/가, -을/를) that change grammatical role but not meaning.
- ▶ Thus, forms like 정부는 / 정부를 all refer to 정부 “government,” yet appear as different items in the corpus.
- ▶ Focusing only on the base form misses occurrences and distorts how a concept (e.g., “government”) is represented.
- ▶ A common solution is to search with a wildcard (e.g., 정부\*), which retrieves all variants (정부가, 정부는, 정부를...).

# Standardising Korean Word Forms & the Role of Concordance Analysis

- ▶ Korean word variants can be normalised by replacing particles/suffixes with a special marker, e.g., converting 북한은 → 북한□, signalling a particle without keeping its form.
- ▶ This prevents tools from treating each variant as a separate item and keeps the core meaning intact.
- ▶ While collocate lists show overall patterns, they cannot reveal contextual meaning.
- ▶ Concordance lines display the search word with surrounding context, helping identify actual usage—e.g., “Great Leader” as a collocate of Kim Jong Il appears in quotation marks, showing it is an official title, not praise.
- ▶ Examining concordances allows researchers to observe recurring patterns, sort contexts, and avoid misinterpreting collocational data.

# Using Concordance Lines to Uncover Discursive Patterns in Corpora

- ▶ Concordance searches reveal both common patterns (e.g., cup of tea in English) and hidden discourses.
- ▶ In North American news, searching North Korea uncovers recurring frames such as “rogue states like NK”.
- ▶ Kim & Choi (2021) found frequent occurrences of 가짜난민 (“fake refugee”) in Korean media, showing negative labelling of refugees.
- ▶ While concordance lines highlight such discursive tendencies, they must be supported with quantitative evidence (e.g., percentages) to ensure objectivity.
- ▶ Context also matters: discourse analysis requires examining longer text passages, not just isolated lines.
- ▶ Even so, concordances provide an essential starting point for identifying patterns that deserve deeper investigation.

# Semantic Prosody & Semantic Preferences in Corpus-Based CDA

- ▶ Traditional CDA faces methodological limitations, but corpus analysis offers more systematic and objective insights.
- ▶ Stubbs (1997) showed how meanings shift over time—e.g., care moving from personal (take care of someone) to institutional (child care).
- ▶ Collocational patterns reveal subtle connotations: little + girl(s) conveys “cute,” while little + man is often pejorative.
- ▶ These patterns relate to semantic prosody—a word’s tendency to occur in positive or negative environments—and to semantic preference, which concerns the semantic categories a word regularly associates with.

- ▶ Zethsen (2006) argued that a word's semantic profile can help analyse social and political perceptions, though such methods remain underused in CDA.
- ▶ The concept originates with Firth (1948) (“phonological colouring beyond semantic boundaries”).
- ▶ Sinclair (1987, 1991) observed that words appear in characteristic semantic environments.
- ▶ Louw (1993) coined semantic prosody as “the constant aura of meaning acquired from collocates”.
- ▶ Partington (1998, 2004) expanded the idea, showing how connotational colouring spreads across collocational patterns rather than individual words.

# The main function of semantic prosody

- ▶ Convey a speaker's or writer's attitude, which can be negative, positive, or neutral (Louw 2000).
- ▶ Scholars use different terms for this attitudinal "aura,"
- ▶ The principal categories are:
  - ▶ negative/unfavourable,
  - ▶ positive/favourable and
  - ▶ neutral.
- ▶ Semantic prosody can be applied in translation studies to reveal dominant discourses.
- ▶ In Korean translations of North Korea-related texts from Newsweek Hangukpan and CNN Hangeul News, a common pattern was 독재자 김정일, showing a clear negative semantic prosody.

# Semantic preference

- ▶ It refers to the tendency of a word (node) to co-occur with semantically related collocates.
- ▶ Oster and van Lawick (2008): describe it as the semantic field that a word's collocates mostly belong to.
- ▶ Semantic preference links a word (node) to a specific set of semantically related collocates, while semantic prosody reflects the broader affective meaning conveyed by a word's typical collocates across wider contexts.
- ▶ For example, words like *happen*, *utterly*, and *cause* have negative semantic prosody because they frequently occur with unpleasant words (utterly pointless, utterly unacceptable), even if those collocates belong to different semantic sets.
- ▶ In contrast, words like dollar, pound, won and money or desk, chair, bed, and table illustrate semantic preference, as they share a clear semantic category.

- ▶ Mautner (2007): showed that corpus analysis is effective for studying discourse and semantic preferences.
- ▶ Using a 57-million-word corpus, she examined the word elderly to uncover age-related stereotypes.
- ▶ She found that its dominant semantic preferences include care, illness, disability, vulnerability, and crime, and that its semantic prosody is strongly negative.
- ▶ Elderly functions less as a neutral age marker and more as a label conveying social assumptions, typically applied to those who are old and dependent, poor or disabled, rather than old and fit, wealthy or powerful.

# Corpus Use in CDA & Semantic Preferences in North Korea Discourse

- ▶ Mautner (2007) supports using corpora in CDA, despite challenges such as limited attention to visual elements.
- ▶ She notes the difficulty of balancing macro vs. micro analysis and quantitative vs. qualitative methods, but argues that corpus approaches help integrate them productively.
- ▶ A manually compiled North Korea corpus revealed semantic preference patterns shaped by the texts' political and social context.

- ▶ In the English North Korea corpus, strong semantic preferences emerge that reflect power dynamics between North Korea and the USA.
- ▶ Phrasal verbs such as condemn, constrain, control, dismantle, deter, halt, rein in, and shut down cluster around discussions of managing North Korea's nuclear activities, typically performed by South Korea, the USA, and bodies like the UN, KEDO, and G8—with the USA present in all actions.
- ▶ Another cluster frames discourse in judicial terms, positioning North Korea as a defendant and the USA as a jury: collocates such as affirm, defend, evidence, experts, investigate, reports, and fact centre on examining North Korea's "illegal" behaviour.

- ▶ Semantic prosody and semantic preference influence each other: semantic preference helps shape prosody, while prosody constrains the choice of collocates (Partington 2004).
- ▶ In translation studies, these patterns can be observed. For example, in a corpus of English news articles and their Korean translations (Newsweek Hangukpan and CNN Hangeul News, 1998–2010), negative semantic prosody surrounds Kim Jong Il, focusing on his appearance, hairstyle, and short height, often mocked in both the source and target texts.
- ▶ Conversely, positive semantic prosody appears around South Korean President Kim Dae Jung, with collocates like engagement, rapprochement, peace, and warm, reflecting his “sunshine policy” promoting reconciliation and cooperation on the Korean peninsula.

- ▶ The study of semantic prosody and semantic preference can reveal shifts in discourse between source texts (ST) and translations (TT).
- ▶ In the English ST corpus, North Korea is portrayed as a voiceless party judged by the USA, reflecting a judicial framing.
- ▶ In the Korean translations, however, North Korea appears to exercise some control through “negotiation” with the USA, showing a subtle shift in discourse.
- ▶ Examining changes in semantic prosodies and preferences thus helps identify dominant discourses in both ST and TT and interpret them in their socio-political context.

# Analytical data: corpus- based CDA that reveals ideological orientations

- ▶ Corpus-based CDA has been widely used in linguistics because concordance lines and collocates allow scholars to identify dominant patterns that construct particular discourses and situate them in socio-political and cultural contexts.
- ▶ This approach has been applied to various genres, including British media discourse on sexism (Pérez-González 2000), government and academic texts on lifelong learning (Piper 2000), and legislative discourse (Graham 2001).
- ▶ Most studies focus on press and newspaper articles, reflecting CDA's aim to reveal the relationship between discourse and hidden power.
- ▶ As Henry and Tator (2002) note, CDA serves as a tool to deconstruct media and elite ideologies and to examine social, economic, and historical power relations between dominant and subordinate groups.

- ▶ Baker and McGlashan (2020): analysed collocates of Romanians in the “Daily Express” and classified them into semantic groups, such as a movement group (arriving, settle, gain) and a law-and-order group (visa, loophole, gang, curbs, controls).
- ▶ Headlines used metaphors like “tide of immigrants” and “swarms of migrants,” figuratively portraying migrants as pests.
- ▶ Distinctively, the study compared patterns in news reports and reader comments.
- ▶ Personal pronouns (they, we, us, them) were more frequent in comments, and verbs linked to us and them (imposing, rob, invade, threaten) constructed a sense of attack or victimhood.
- ▶ The study argues that articles and comments operate together, shaping public perception—such as influencing the Brexit decision—and demonstrates that news corpora can uncover underlying media ideologies and attitudes toward particular groups.

# Corpus-Based CDA of Terrorism Discourse: Qian (2010)

- ▶ Qian (2010) conducted a corpus-based CDA study on terrorism, analysing People's Daily (China) and The Sun (UK) before and after 9/11 (2000–2002).
- ▶ The study examined how discourses about terrorism shifted over time in both national contexts.
- ▶ Although not a translation study, it provides valuable insights because the two newspapers differ significantly:
  - ❖ The Sun is a UK tabloid, operating in a system without government-controlled media.
  - ❖ People's Daily is a Chinese broadsheet, published within a state-censored media environment.

- ▶ Corpus-based approaches have been applied to study translation shifts in ideology and register.
- ▶ Mouka, Saridakis, and Fotopoulous (2015): examined English STs and their Spanish and Greek translations in films about racism (1989–2006), using appraisal theory to analyze negative expressions.
- ▶ They found that translators sometimes mitigated racial slurs, but in other cases intensified racist attitudes.
- ▶ Similarly, Li and Pan (2021): analyzed Chinese political texts and their English translations, revealing that individual translators sometimes diverge from official ideological norms, adopting target-oriented strategies that alter the intended discourse, despite established government guidelines.

# Corpus- based CDA approach to translation research

- ▶ A corpus-based CDA approach is valuable for translation research. House (2015) highlights that corpora help assess whether a translation aligns with target-culture genre norms.
- ▶ In corpus-based CDA of translations, the focus is often on optional shifts—changes made by translators for stylistic, ideological or cultural reasons.
- ▶ The analysis typically involves four steps:
  - 1) identifying recurrent patterns in both the ST and translation corpus,
  - 2) comparing these patterns,
  - 3) examining longer stretches of text in detail, and
  - 4) interpreting any shifts or deviations in relation to the historical, political, and social contexts of the ST and TT, using relevant analytical frameworks.

- ▶ A translational corpus can include multiple sub-corpora from different media outlets, enabling more comprehensive analysis.
- ▶ The analysis of a translational corpus can focus on three aspects:
  - 1) patterns in the English ST and Korean TT,
  - 2) recurrent patterns specific to a media outlet, and
  - 3) shifts between the ST and TT.
- ▶ Such shifts can then be interpreted within their socio-political and historical contexts. For example, in a corpus of Newsweek and CNN articles about North Korea (1998–2010) and their Korean translations, the English ST frequently used leader to refer to Kim Jong Il, whereas the Korean TT disproportionately employed the formal designation Chairman/Chairman of the National Defence Commission.

- ▶ Translational corpus analysis can examine:
  - 1) patterns in the English ST and Korean TT,
  - 2) outlet-specific recurrent patterns, and
  - 3) shifts between ST and TT, which are interpreted in their socio-political and historical contexts.
  
- ▶ For instance, in Newsweek and CNN articles on North Korea (1998–2010) and their Korean translations, leader commonly referred to Kim Jong Il in the ST, while the TT predominantly used the formal title Chairman of the National Defence Commission. The Joong Ang Ilbo (1997) guidelines for Newsweek Hangukpan allowed historical figures or well-known names to be referred to without full titles, e.g., Churchill instead of Churchill, the then prime minister.

- ▶ Similarly, Kim Jong Il could appear without his official title (Chairman of the National Defence Commission), though Chairman was the default in Korean translations, even when the English ST frequently used leader.
- ▶ This demonstrates translator agency and may reflect the political climate at the time.
- ▶ Such choices suggest that, for a South Korean audience, North Korea was framed more as a neighbouring country than as a consistently provocative “enemy regime,” illustrating how minor cumulative shifts in translation can influence discourse (Bielsa and Bassnett 2009).

- ▶ Corpus-based CDA is increasingly used to examine optional translation shifts in translation and interpreting, contextualized within historical and socio-political settings.
- ▶ For example, Li and Zhang (2021) built a corpus of interpreting transcripts from Chinese Foreign Minister Wang Yi's press conferences (2016–2018) to study the gatekeeping role of interpreters.
- ▶ Using a CDA framework, they analyzed shifts in interpersonal metafunctions—mood, speech function and appraisal—and found that interpreters often withhold, modify or supplement messages, especially when responding to foreign journalists' sensitive questions.
- ▶ These shifts reflect ideological and institutional alignment with the Chinese government.

- ▶ Corpus-based, CDA-informed translation analysis is a key research area for Tao Li, who combines corpus tools with van Dijk's ideological square model to identify linguistic patterns and translation shifts (Li and Zhu 2020; Li and Pan 2021).
- ▶ For example, Ping (2021) analyzed a corpus of translated media coverage of the 2014 Hong Kong protests from Reference News (China), EJ Insight (Hong Kong), BBC Chinese (UK), and NYT Chinese (US) using Sketch Engine.
- ▶ By examining high-frequency words and semantic prosodies, Ping found that BBC and NYT Chinese linked the protests strongly with democracy, whereas Reference News downplayed democratic associations and included less favorable connotations, revealing discursive differences in translation.

- ▶ They proposed a typology of “manipulative operations” in translation:
  - 1) blocking undesired ST representations,
  - 2) demoting undesired ST representations,
  - 3) promoting desired ST representations, and
  - 4) preserving desired ST representations, using techniques like omission, euphemism, modality changes, addition, and lexicalization.
- ▶
- ▶ These studies show that translation shifts, whether conscious or unconscious, align with the socio-political expectations of the target audience.

Thank You ...