# Contrastive Analysis, *Tertium comparationis,* and Corpora

## Signe Oksefjell Ebeling & Jarle Ebeling

# Outline

- Doing contrastive analysis
- Types of corpora: their pros and cons
- Case study
  - Based on comparable data only
  - Based on bidirectional parallel data

# Our background/position

- Corpus-based contrastive analysis
- Starting point: perceived (dis)similarities
- Bi-/multidirectional, parallel corpora
- Translations seen as the best *tertium comparationis*
- Interested in finding systemic differences / differences at the system level of languages, i.e. not primarily motivated by applied research

Inspired by James (1980), Ivir (1983, 1990), Johansson (1994 [& Hofland], 2007), Chesterman (1998, 2007), Altenberg (1999)

Awareness of research in neighbouring fields, e.g. descriptive translation studies and typology

# Starting point of the CA

1. Chesterman (1998; 2007)

a) Perceived similarity of any kind between phenomenon X in language A and phenomenon Y in language B (cf. Chesterman 1998)

b) What is the nature of the similarity (form, meaning, function)?

c) Describe the relationship between X and Y in the compared languages; or as is more often the case, the relationship between X in language A and $Y_1$, $Y_2$, $Y_3$, etc. in language B

d) Use the description to enrich knowledge of the individual languages and/or the relationship between the languages compared

# Starting point of the CA

2.  Perceived (quantitative) dissimilarities/ difference between original and translated text in the same language

3.  Frequent omissions and/or additions in the translation

4.  Exploratory by e.g. starting from a unit or item (e.g. word, frame, pattern, construction) in one language and classify the correspondences in one or more other languages

Close, qualitative scrutiny of the differences is necessary.

## Corpora for contrastive analysis

**Comparable**
(period / genre / domain)

**Parallel**
(topic, function, meaning, style)

Other types of corpora used in (other types of) contrastive analysis
- Comparative corpora (regional, dialectal comparisons; also diachronic studies of same language (family)?)
- Translation-only corpora (mono-, bi- or multilingual)
- Multimodal (gestures, images, descriptions/interpretation of situations)
- …

# Pros and cons of corpus types for CA

| | | Pros | Cons |
|---|---|---|---|
| Comparable corpora | | • Not restricted to translated text types<br>• More readily available<br>• Comparison of orig. language | • Criteria for comparability / **TC**?<br>• No alignment possible<br>• Cannot reveal sets of cross-linguistic correspondences |
| Parallel corpora | Unidirectional translation corpora | • Alignment is possible<br>• Meaning & function constant across the languages (**a relatively sound TC is present**)<br>• Possible to discover (sets of) cross-linguistic correspondences ('translation paradigms') | • Restricted range of text types<br>• Translation effects:<br>  • (i) Traces of source language in translated texts<br>  • (ii) Traces of the translation process, including errors |
| | Balanced, bi-/ multidirectional translation corpora | • Same as unidirectional translation corpora PLUS:<br>• Possible to check translations in both directions (control for translation effects)<br>• Comparison of orig. language<br>• **A sounder TC is present** | • Even more restricted range of text types<br>• Achieving balance of text types between directions of translations |

Cf. Altenberg & Granger (2002), Johansson (2007), Aijmer (2008), Ebeling & Ebeling (2013), Hasselgård (2017)

# Short case study

- **Starting point:**
  – Previous contrastive study of two similar-looking patterns in English and Norwegian: *for * sake/for * skyld* (* = genitive)

  (Ebeling & Ebeling 2014)

- **Today's experiment:**
  – Carve up previous study in a different way to demonstrate different TCs:
    - Sameness of form in a comparable corpus;
    - Translation correspondence in a bidirectional parallel corpus.

- **The corpora:**
  – The English and Norwegian original, comparable texts of the English-Norwegian Parallel Corpus+;
  – The English and Norwegian original and translated, comparable texts of the English-Norwegian Parallel Corpus+

# Preliminaries
## Perceived similarity

**for the <span style="color:red">purpose</span> of (something)**

(1)    We stayed together <u>for appearances' sake</u>, … (PeRo2E)

(2)    … det videre søket etter Leike ville bare være <u>for syns skyld</u>. (JoNe2N)

**out of <span style="color:red">consideration</span> for or in order to help or please someone**

(3)    Just for a while, then, another chapter or two — <u>for Miriam's sake</u>. (PaAu1E)

(4)    … men <u>for Mathias' skyld</u>, herregud, de hadde jo et barn sammen !  (JoNe1N)

**to express impatience, annoyance, urgency, or desperation (<span style="color:red">expletive</span> use)**

(5)    <u>For God's sake</u>, stop! (MiWa1E)

(6)    Si noe, <u>for Guds skyld</u>! (LSC2)

(*Oxford Dictionaries Online* and Altenberg 1982)

# Comparable version of study

- Preferred use in English originals:
  – Expletive>Consideration>Purpose

- Preferred use in Norwegian originals:
  – Purpose>Consideration>Expletive

# Summary of the comparable study
## with ref. to Sinclair's (1996) Extended-units-of-meaning model

| | English (Core: *for * sake*) | | Norwegian (Core: *for * skyld*) |
|---|---|---|---|
| Collocation | for [*God's / Christ's / heaven's*] sake | Collocation | for [*sikkerhets*] skyld |

How to get at the closest correspondences in the other language?

- E.g.: How is the expletive use of *for * sake* typically expressed in Norwegian if not with the formally similar pattern?

| | | | |
|---|---|---|---|
| preference | sex | preference | arrangement (order, safety, simplicity) |
| Semantic prosody | Annoyance (shown by words such as *God's* (religion) / *fuck's* (sex) or "pretend swear words" such as *goodness'*) | Semantic prosody | Purpose (shown by words such as *sikkerhets* (*'safety's'*) / *enkelhets* (*'simplicity's'*) |

# Bidirectional version of study

Translations of *for * sake* and *for * skyld* in the ENPC+ (raw numbers)

| Congruent / non-congruent | EO > NT | NO > ET |
|---|:---:|:---:|
| *for * sake = for * skyld* | 27 | 18 |
| *for * sake/skyld* = 'other' | 58 | 46 |
| Total | 85 | 64 |

**Congruent** = formally similar correspondences (i.e. *for * sake = for * skyld*)

**Non-congruent** = formally dissimilar correspondences (e.g. *for God's sake* = *Herregud* ('Lordgod')

Mutual correspondence (Altenberg 1999) of *for * sake* and *for * skyld* = 30.2%

## **Examples of main tendencies**

English expletive → non-congruent

(7)     "They're just traffic cones, <u>for fuck's sake</u>." (PeRo2E)

"Det er jo bare trafikkjegler, <u>for faen</u>." (PeRo2TN)

Lit.: … for the devil

English consideration → congruent

(8)     <u>For my sake</u>. (JB1)

<u>For min skyld</u>. (JB1T)

English purpose → congruent

(9)     We stayed together <u>for appearances' sake </u>… (PeRo2E)

Vi holdt sammen <u>for syns skyld </u>… (PeRo2TN)

## Norwegian expletive → congruent

(10)   - Slipp han nå <u>for guds skyld </u>ned. (PePe1N)

"Put him down, <u>for God's sake</u>." (PePe1TE)

## Norwegian consideration → congruent

(11)   … hun gjorde det <u>for pappas skyld </u>… (PeRy1N)

… she was doing it <u>for Dad's sake </u>… (PeRy1TE)

## Norwegian purpose → non-congruent

(12)   Han rygget et skritt <u>for sikkerhets skyld</u>. (KaFo1N)

Lit.: … for safety's sake

He retreated a step, <u>just to be on the safe side</u>. (KaFo1TE)

## Non-congruent English correspondences (translations and sources) of the most frequent Norwegian purpose uses with *for * skyld*

| *for sikkerhets skyld* ('for safety's sake') | *for moro skyld* ('for fun's sake') | other |
|---|---|---|
| to be on the safe side | | |
| just in case | | |
| | for N (fun/entertainment/pleasure) | |
| | | a matter of N |
| | | for the sake of N |

## Non-congruent Norwegian correspondences (translations and sources) of the most frequent English expletives with *for * sake*

| for Christ's sake | for fuck's sake | for God's sake |
|---|---|---|
| *for svingende* (lit. 'for swinging') | | *for svingende* (lit. 'for swinging') |
| *herregud* ('God', lit. 'lordgod') | | *herregud* ('God', lit. 'lordgod') |
| helvete ('hell') | | *i herrens navn* ('in the lord's name') |
| | (*for*) *faen* ('(for) the devil') | (*for*) *faen* ('(for) the devil') |

**Summary of the bidirectional parallel study**

- Uncover different preferences of use in English and Norwegian (incl. different extended-units-of-meaning) through both comparable and translation data.

- Translation as a TC offered corresponding, and arguably more equivalent, expressions of *for * sake* / *for * skyld* in the other language.

  – E.g. typical Norwegian expletives corresponding to the expletive *for * sake* pattern emerged.
  – E.g. typical English expressions of purpose corresponding to the purpose *for * skyld* pattern emerged.

- The patterns are cross-linguistically most similar to each other in the consideration use.

- Gained a deeper understanding of the pattens compared.

# Summing up: CA, TC and corpora

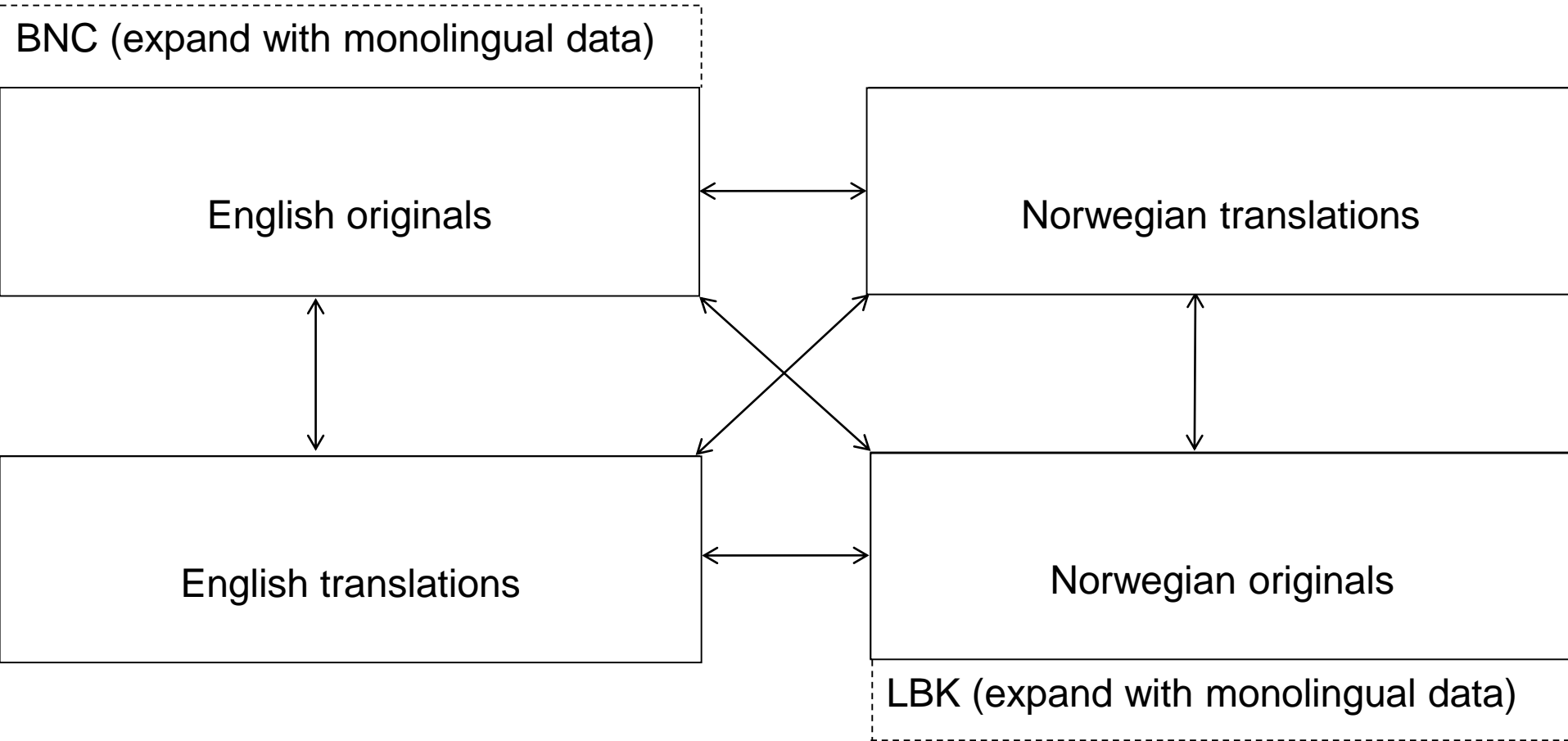| Comparable only | Parallel (bidirectional)<br>**(incl. omissions/additions in translated text)** |
|---|---|
| CA based on researcher's bilingual knowledge, dictionaries, grammars:<br><br>• Perceived (dis)similarity<br> • predefined items/ categories | CA based on researcher's bilingual knowledge, dictionaries, grammars, and **translators' bilingual knowledge**:<br><br>• Perceived (dis)similarity<br> • predefined items/ categories<br><br>CA based (primarily) on **translators' bilingual knowledge/competence**:<br><br>• Exploratory: predifined item with translation paradigm<br><br>• Exploratory: frequency differences in original vs. translated texts (i.e. undefined starting point) |

**Point for discussion**

Are there other/better ways of doing corpus-based CA (regardless of object of study)?

# References

Aijmer, Karin. 2008. Parallel Corpora and Comparable Corpora. In Lüdeling, A. & M. Kytö (eds.), *Corpus Linguistics. An International Handbook, Vol. 1.* Berlin / New York: Walter de Gruyter. 275-292.

Altenberg, B. 1982. *The Genitive v. the of-construction: A Study of Syntactic Variation in 17th Century English*. Lund: Gleerup.

Altenberg, Bengt. 1999. Adverbial connectors in English and Swedish: Semantic and lexical correspondences. In Hasselgård & Oksefjell (eds), 249–268.

Altenberg, Bengt & Granger, Sylviane (eds). 2002. *Lexis in Contrast. Corpus-based Approaches* [Studies in Corpus Linguistics 7]. Amsterdam: John Benjamins.

Chesterman, Andrew. 1998. *Contrastive Functional Analysis* [Pragmatics & Beyond New Series 47]. Amsterdam: John Benjamins.

Chesterman, Andrew. 2007. Similarity analysis and the translation profile. In The Study of Language and Translation, Willy Vandeweghe, Sonia Vandepitte & Marc van de Velde (eds). *Belgian Journal of Linguistics* 21: 53–66.

Ebeling, Jarle & Ebeling, Signe Oksefjell. 2013. *Patterns in Contrast*. Amsterdam: John Benjamins.

Ebeling, Signe Oksefjell & Ebeling, Jarle. 2014. *For Pete's sake*! A corpus-based contrastive study of the English/Norwegian patterns "for * sake" /" for * skyld". *Languages in Contrast 14*(2), 191-213.

Hasselgård, Hilde. 2017. Corpus-based contrastive analysis: some methodological issues and a talking case study. Guest lecture given on 5 December 2017, Charles University, Prague.

Ivir, Vladimir. 1983. A translation-based model of contrastive analysis. In *Cross-language Analysis and Second Language Acquisition* 1 [Jyväskylä Cross-language Studies 9], Kari Sajavaara (ed.), 171–178. Jyväskylä: Department of English, University of Jyväskylä.

Ivir, Vladimir. 1990. Languages in contact through translation. In *Proceedings of the Symposium 16.1 Languages in Contact of the 12th International Congress of Anthropological and Ethnological Sciences* – Zagreb, July 25–27, 1988, 88–92. Zagreb: Institute of linguistics, University of Zagreb.

James, Carl. 1980. *Contrastive Analysis*. London: Longman.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins.
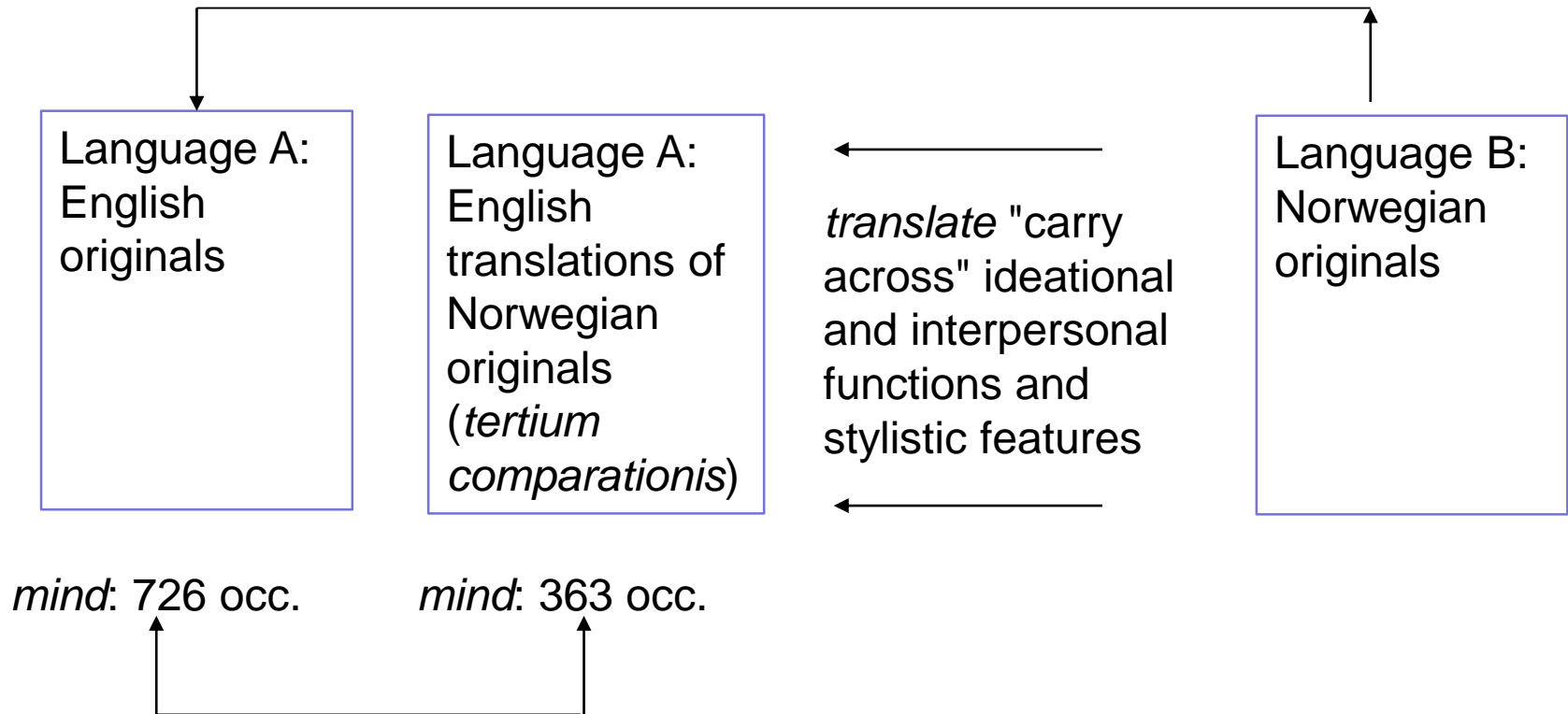
# References cont.

Johansson, Stig & Hofland, Knut. 1994. Towards an English-Norwegian Parallel Corpus. In *Creating and Using English Language Corpora: Papers from the Fourteenth International Conference on English Language Research on Computerized Corpora*, Zurich 1993, Udo Fries, Gunnel Tottie & Peter Schneider (eds), 25–37. Amsterdam: Rodopi.

*Oxford Dictionaries Online* (https://www.oxforddictionaries.com/)

Sinclair, John. 1996. The search for units of meaning. *Textus* IX: 75–106.

# The structure of the ENPC+

BNC (expand with monolingual data)

English originals

Norwegian translations

English translations

Norwegian originals

LBK (expand with monolingual data)

The corpus "permits contrastive studies […] based on original texts and translations as well as on parallel original texts" (Johansson & Hofland 1994: 27), i.e. the corpus can be seen both as a bidirectional translation corpus and a comparable corpus.

# Un-/discover differences and similarities

How to discover and investigate lexical gaps, e.g. Norw. 'jo' (*but, just, only, in fact, of course, after all, Ø*)

| Language A: English originals | Language A: English translations of Norwegian originals (*tertium comparationis*) | *translate* "carry across" ideational and interpersonal functions and stylistic features | Language B: Norwegian originals |
|---|---|---|---|

*mind*: 726 occ.          *mind*: 363 occ.

Quantitative differences in a
balanced corpus: Reason/Cause?