



UiO : **University of Oslo**

Hilde Hasselgård

Idiomaticity in English and Norwegian:
combining contrastive and interlanguage
corpus studies



The idiomaticity project

- The idiomaticity project combines corpus-based contrastive analysis and learner corpus research.
- Focus on English-Norwegian contrasts in the interface between lexis and grammar.
- A particular interest in the idiomatic, nativelike, (co-)selection of words and phrases.

Backdrop

- Monolingual studies, particularly of English, have demonstrated that both lexis and grammar vary across registers, disciplines and genres:
→ expanding the focus to a multilingual context.
- Languages differ in the ways recurrent word combinations are composed and in their preferred (types of) phraseological patterns (e.g. Ebeling & Ebeling 2013, Granger 2014)
- Fløttum et al (2006) have demonstrated how certain discourse features vary across both languages and academic disciplines.

Research areas of the Idiomaticity project

- i. how patterns vary across languages;
- ii. how patterns vary across registers;
- iii. whether register features are similar or different cross-linguistically;
- iv. how patterns vary across English as a first (L1) and second language (L2);
- v. the extent to which learners of English manage to adapt their lexicogrammatical patterns to the appropriate register characteristics of English L1;
- vi. the extent to which cross-linguistic differences are reflected in the English of advanced learners;
- vii. how lexicogrammatical patterns in different languages reflect cultural differences;
- viii. the extent to which learners apply the cultural patterns of their first or their second language.

Long-term linguistic and methodological aims

- **Patterns of lexis and grammar:** to identify salient lexico-grammatical patterns, and their structural and functional features, in the individual registers and languages represented in our corpora.
- **Lexicogrammatical register and discipline features:** to identify and explain variation in lexicogrammatical patterns across registers and disciplines.
- **Cultural and contextual features:** to interpret variation in lexicogrammatical patterns in the light of national and disciplinary cultures.
- **Method for contrastive analysis:** to refine the procedures for conducting contrastive analysis based on comparable corpora, to increase the reliability of the comparison.

Corpora

- Available corpora:
 - The English-Norwegian Parallel Corpus
 - International Corpus of Learner English, VESPA, LINDSEI
 - The KIAP corpus (linguistics, medicine, economics; English, Norwegian, French)
 - L1 reference corpora in English and Norwegian
- Desirable corpus development:
 - A larger ENPC, particularly adding more non-fiction
 - More disciplines (and more texts) in VESPA
 - Literature component to complement KIAP
 - Monolingual novice corpus in Norwegian L1

Translation corpora, comparable corpora

- The ideal solution for contrastive studies: bidirectional translation corpora.
 - Challenge: not all types of text are translated
- Main problems of comparable corpora (similar originals in two languages): identifying a *tertium comparationis*; knowing what to compare.
- Exploiting the benefits of a bidirectional corpus in comparable corpus studies:
 - Using the two corpus types in conjunction (in spite of genre/register differences) to identify the TC.

Expanding the fields of CA and CIA to phraseology

Methodological / practical challenges

- Extraction methods – corpus-driven or corpus-based?
 - Simplexes with context
 - N-grams
 - Phrase frames (colligational frameworks)

Extraction methods and comparisons made in contrastive phraseological investigations

- Johansson 1998/2007: *love/hate*; 2009 (*way*; *V-ing one's way*)
- S. Ebeling 2014 – cross-linguistic phraseology of *eye/øye*
- Hasselgård (forthcoming): lexical bundles containing the predefined node words *place* and *plass/sted*.

- Ebeling et al 2013: n-gram-based comparison of originals and translations within the same language
- Ebeling & Ebeling 2013: n-gram-based starting point, with additional criteria of semantic/syntactic unity for contrastive (bidirectional) analysis
- Hasselgård (2017) and Ebeling & Ebeling (2017): attempt at direct cross-linguistic comparison of n-grams shown to be problematic

- Hasselgård (2016): colligational framework (the N1 of the N2) – works only in one direction of comparison.

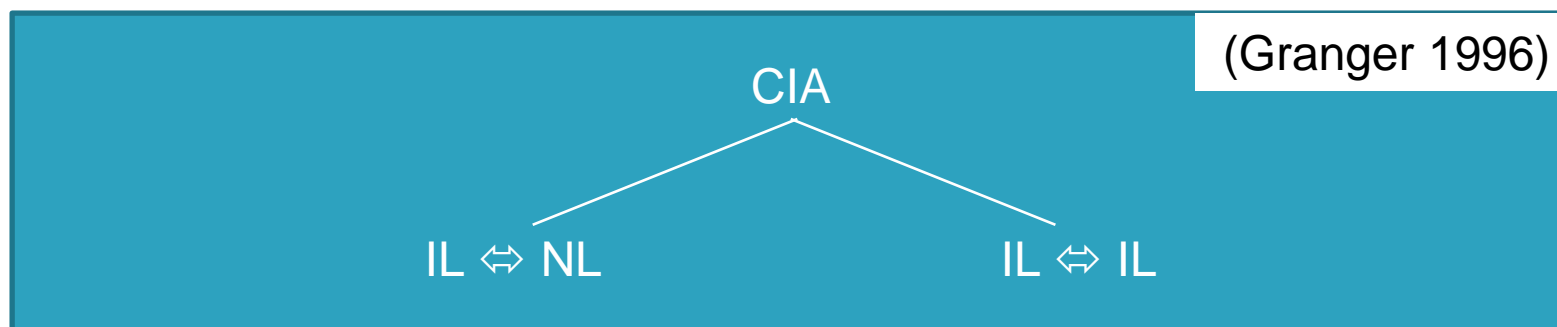
Challenges of (phraseological) CA/LCR (cont.)

- Systemic differences between languages
 - E.g. V2 vs SV word order; article or suffix to mark definiteness of nouns:
I understand now / Now I understand – jeg forstår nå / Nå forstår jeg
in the city – i byen
- Translation corpora: problem of limited size and genre match with learner corpora
- Corpus comparability: parallel vs. learner corpora; Learner corpora vs. L1 reference; genre/register across languages and cultures.
- L1 and L2 varieties (or source/target varieties) of the same language: how to gauge cross-linguistic influence (transfer / translation effects)

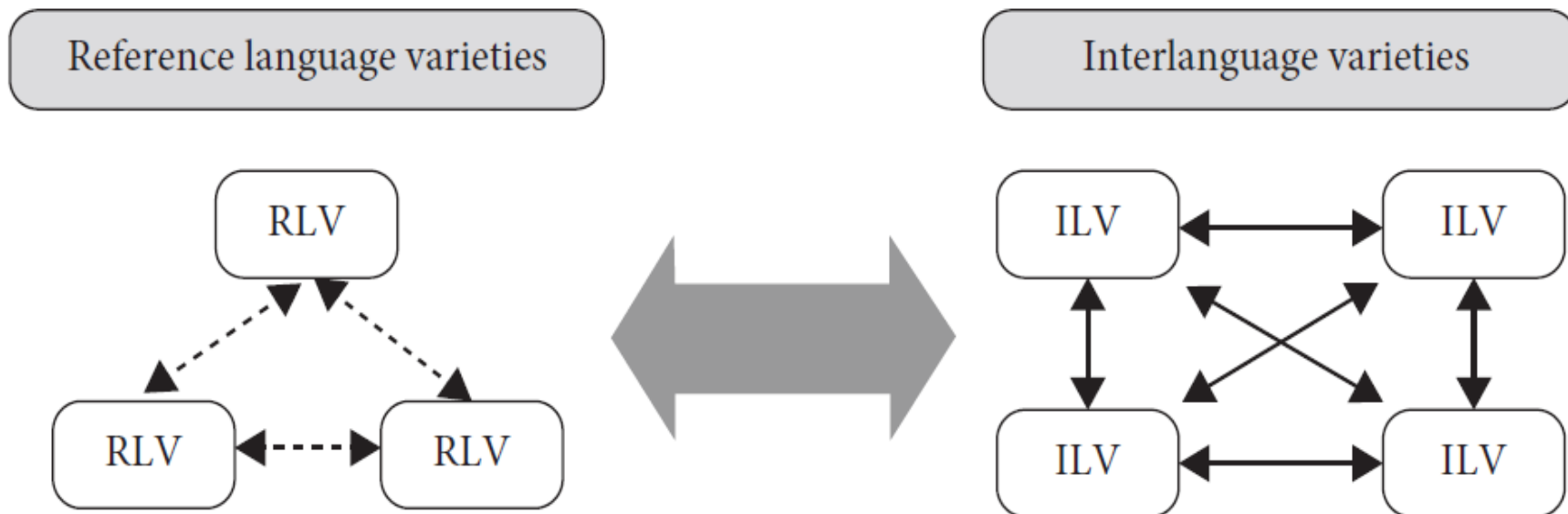
Contrastive Interlanguage Studies (CIA)

- Comparing a non-native variety – interlanguage/ L2 – to a relevant L1 variety.
- Comparing different non-native varieties (e.g. Norwegian and French learners of English)
- CIA studies typically aim to discover areas of quantitative and qualitative difference between L1 and L2 production and/or areas of transfer from the learners' L1.

Contrastive Interlanguage Analysis and learner corpus research



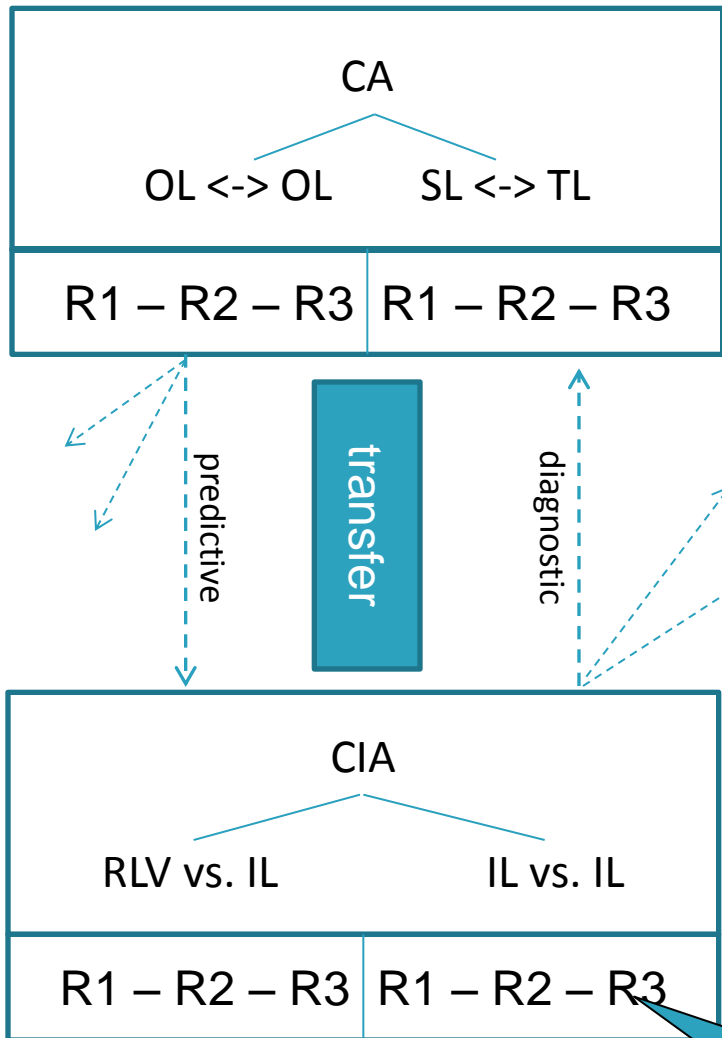
CIA² (Granger 2015)



Combining contrastive analysis and studies of learner language

- The idea that the two fields are connected is not new, e.g.
 - Fries (1945, 9): “the most effective materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner”.
 - Lado (1957, 2): “the teacher who has made a comparison of the foreign language with the native language of the students will know better what the real learning problems are and can better provide for teaching them”.
- The relationship is not straightforward:
 - not all differences between a learner’s L1 and the L2 cause errors or learning difficulties (Gilquin 2000/2001, 101)
 - *perceived* language distance can act as a constraint on negative transfer (R. Ellis 2008, 397).
 - if learners do not perceive or assume any degree of similarity between their L1 and the L2, transfer does not take place (Ringbom 2007, 26).

**The Integrated Contrastive Model
(adapted from Granger 1996,
Gilquin 2000/2001)**



Key:
 CA = contrastive analysis;
 OL = original language;
 SL/TL= source/translated language;
 R= register;
 CIA = contrastive interlanguage analysis;
 RLV = reference language variety;
 IL = interlanguage (second language)

New! Added register

Predictive direction

- Identify point(s) of cross-linguistic difference (or similarity)
 - Qualitative/quantitative
- Hypothesize effect (ease or difficulty) in L2
- Check hypothesis

- Example: Johansson (2007) found that *seem* is more frequent than its Norwegian copular and catenative equivalents (*The idea seems acceptable / The idea seems to be acceptable*). The Norwegian correspondences often involved an experiencer (e.g. *to me*).
- Hypothesis: Norwegian learners will underuse *seem*, but the experiencer role will be overrepresented.
- Findings of the CIA (ICLE/LOCNESS): Norwegian learners do *not* underuse *seem* – they overuse it. Hypothesis rejected.
- Explicit experiencers are twice as common in ICLE as in LOCNESS: hypothesis confirmed.
- Results tie in with hedging and writer/reader visibility.

Diagnostic direction

- Identify discrepancy (or striking similarity) between IL and comparable reference language variety (L1).
 - Qualitative/quantitative
- Seek explanation in contrastive (cross-linguistic) analysis.
- Example: Aijmer (2002) finds that Swedish learners use modal combinations (e.g. *may perhaps, maybe would*) more frequently than native speakers. She explains this by recourse to contrastive analysis: such combinations are more common in Swedish than in English (ibid., 69).
 - *Vi kan kanskje sammenligne med en kunstmaler.* (JG1)
[Lit: *We can perhaps compare with a painter*]
We can make a comparison to painting. (JG1T)

An “in the case of” study to illustrate the model

- A relatively frequent 4-gram in English (non-fiction) with no clear Norwegian counterpart.
- Raw frequencies in the ENPC:

<i>in the case of</i>	Original	Translation
Fiction	1	1
Non-fiction	23	6
Total	24	7

- A register effect
- A translation effect

Non-fiction: Recurrent Norwegian correspondences of *in the case of*

Translations (23)		Sources (6)	
for	7	ved	4
når det gjelder	4	Other	2
(slik) tilfellet er/var	3		
i tilfellet med	2		
Other	7		

Fiction: both examples of *in the case of* correspond to *når det gjelder*

Når det gjelder selected for further analysis because of frequency plus functional (and partly constructional) similarity.

Når det gjelder – correspondences in ENPC non-fiction (raw frequencies)

Translations (38)		Sources (57)	
in terms of	6	in	11
with regard to	6	as regards	6
when it comes to	3	of	6
in (3)	3	in terms of	4
where X is concerned	2	in the case of	4
∅	10	for	4
Other	8	about	3
		at	2
		concerning	2
		regarding	2
		with regard to	2
		∅	6
		Other	5

Mutual correspondence of *in the case of* and *når det gjelder* (non-fiction)

- English source → Norwegian translation: 17.4%
- Norwegian source → English translation: 0%
- Mutual correspondence: 6.6%

- Reverse mutual correspondence: 6.3%
 - Also asymmetrical, with 0% in N→E.

I.e. no clear Norwegian source of *in the case of* → hypothesis that Norwegian learners will underuse the expression.

In the case of register (BNC)

SECTION	FREQ	SIZE (M)	PER MIL
SPOKEN	111	10.0	11.14
FICTION	91	15.9	5.72
MAGAZINE	173	7.3	23.82
NEWSPAPER	141	10.5	13.47
NON-ACAD	859	16.5	52.08
ACADEMIC	2,043	15.3	133.25
MISC	1,211	20.8	58.12
TOTAL	4,629		

In the case of L1 vs L2

		N	Per 100,000 w
Learner varieties	ICLE-NO	4	1.9
	VESPA-NO (ling)	28	10.5
Reference varieties (L1)	LOCNESS	28	8.6
	BAWE (ling)	29	17.3
	KIAP (ling)	69	15.7

- L1/L2 difference – underrepresentation in L2
- Register difference – general argumentative vs. disciplinary writing – visible in L2 in spite of general underrepresentation

Conclusions in the case of *in the case of*

- The English expression does not have a formally similar correspondence in Norwegian.
- (r)MC values between *in the case of* and functionally similar Norwegian correspondences are low (and asymmetrical).
- Norwegian learners may avoid/underuse the expression for these reasons.
- *In the case of* is register-sensitive, typical of academic writing.
- Norwegian learners, contrary to expectation, seem to have grasped the register difference and use the expression more in disciplinary writing than in general argumentative texts (but underuse it in both registers).
- In spite of the low MC values, the translation paradigms of *in the case of* and *når det gjelder* provide a basis for exploring functionally equivalent expressions cross-linguistically in comparable corpora.

An area for further study

- It appears that Norwegian learners of English – possibly on the basis of preferences in Norwegian L1 – choose the pattern “in the N of” less often than English L1 users (tendency in both ICLE/LOCNESS and VESPA/BAWE).
 - On the other hand, Norwegian learners are fond of more verbal expressions such as *when it comes to*.
- I.e. a possible contrastive difference in preferences of (complex) nominals vs. verbal expressions (needs further contrastive study)

Other types of implications from register-based ICM studies

- Cross-cultural rhetoric
 - E.g. signs of more writer-/reader-orientation (interactive features) in Norwegian academic writing and in Norwegian-produced English than in English L1 (Fløttum et al 2006, Paquot et al 2013, Hasselgård forthcoming)
 - Exploring whether/how “corpora provide a method of observing how recurrent phraseology categorizes events and therefore expresses cultural knowledge” (Stubbs 2014)

UiO : University of Oslo

- Aijmer, K. 2002. Modality in Advanced Learners' Written Interlanguage. In S. Granger, J. Hung and S. Petch-Tyson (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 55–76. Amsterdam: Benjamins.
- Ebeling, J. & S. O. Ebeling. 2013. *Patterns in Contrast*. Amsterdam/Philadelphia: Benjamins.
- Ebeling, J., S. O. Ebeling & H. Hasselgård. 2013. Using recurrent word-combinations to explore cross-linguistic differences. In Aijmer, K. and B. Altenberg (eds), *Advances in Corpus-based Contrastive Linguistics. Studies in honour of Stig Johansson*. Amsterdam: Benjamins, 177-199.
- Ebeling, S. O. 2011. Recurrent word-combinations in English student essays. *Nordic Journal of English Studies* 10(1), 49-76.
- Ebeling, S. O. 2014. An eye for an eye? Exploring the cross-linguistic phraseology of *eye/øye*. *Nordic Journal of Linguistics*. 37(2), 225-255.
- Ebeling, S.O. and H. Hasselgård. 2015. Learners' and native speakers' use of recurrent word-combinations across disciplines. In *Bergen Language and Linguistics Studies (BeLLS)* vol. 6, 87-106.
- Gilquin, G. 2000/2001. The Integrated Contrastive Model: Spicing up your data. *Languages in Contrast* 3(1): 95-124.
- Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, B. & M. Johansson, M. (eds), *Languages in Contrast. Papers from a symposium on text-based cross-linguistic studies, Lund 4-5 March 1994*. Lund: Lund University Press, 37-51.
- Granger, S. 2014. A lexical bundle approach to comparing languages: stems in English and French. *Languages in Contrast*, 14(1), 58-72.
- Hasselgård, H. 2016. *The way of the world*: The colligational framework “the N1 of the N2” and its Norwegian correspondences. *Nordic Journal of English Studies* 15:3, 55-79.
- Hasselgård, H. Forthc. Phraseological teddy bears: frequent lexical bundles in academic writing by Norwegian learners and native speakers of English. In M. Mahlberg and V. Wiegand (eds), *Corpus Linguistics, Context and Culture*. Berlin: De Gruyter Mouton.
- Hasselgård, H. Forthcoming. Lexical patterns of PLACE in English and Norwegian. To appear in H. Dirdal and T. Egan (eds), *Cross-linguistic Correspondences: From Lexis to Genre*. Benjamins.
- Hasselgård, H. & S. Johansson. 2011. Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds), *A Taste for Corpora. In honour of Sylviane Granger*. Amsterdam: Benjamins, 33-62.
- Johansson, S. 2009. Which way? On English *way* and its translations. *International Journal of Translation* 21(1-2), 15-40
- Johansson, S. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: Benjamins.
- Pawley, A. & F. H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In *Language and communication*, eds. J. C. Richards and R. W. Schmidt, 191–226. London: Longman.
- Stubbs, M. 2014. Searle and Sinclair on communicative acts. A sketch of a research problem. In M. Gómez González, F. J.R.dM. Ibáñez, F. González García, A. Downing (eds), *The Functional Perspective on Language and Discourse. Applications and Implications*. Amsterdam: Benjamins, 243-260.