Contrastive Phraseology: Method and Analysis

Signe Oksefjell Ebeling
s.o.ebeling@ilos.uio.no

Prague, 27 March 2018
[T]he corpus revolution has brought the contextual patterning of words into focus in recent years. The use of idioms and collocations tends to be highly language-specific and the syntagmatic aspect of lexis is consequently of great contrastive interest, theoretically as well as practically.

(Altenberg & Granger 2002: 30)
Outline

• Preliminaries
  – CA on the basis of bidirectional translation data
  – Phraseology
• Patterns and extended units of meaning
• Method applied in the analysis of patterns in contrast
• Case study: *out of the ordinary*
• Concluding remarks
CA on the basis of translations

- Translation correspondence as
  - perceived similarity (Chesterman 1998, 2007);
  - measure of equivalence; mutual correspondence measures the intertranslatability of items (Altenberg 1999)
- Johansson’s (2007) bidirectional parallel corpus model applied to extraction of n-grams

The corpus can be said to lend an element of empirical inter-subjectivity to the concept of equivalence, especially if the corpus represents a variety of translators.

(Altenberg & Granger 2002: 17)
Schematic layout of the procedure of doing contrastive analysis based on parallel corpora, containing translations (Ebeling & Ebeling 2017: 33)
Guilt brings us nearer to God.

Bring in something to eat …

…to bring me a bottle of …

…what brings you here?

…to bring you luck, and it has ...

…and bring them back.

…to bring her home, ...

…papers report.

…skulle bringe deg lykke, og det har ...

…og bringe dem tilbake.

…for å bringe henne hjem ...

…avisene bringer.
Framework of correspondence

Ebeling (2017); based on Johansson (2007: 25)
Correspondences

Overt congruent translation
(1) Guilt brings us nearer to God. (FW1)
   Skyldfølelse bringer oss nærmere Gud. (FW1T)

Overt non-congruent translation
(2) Aila understood everything, even the things he didn't intend to bring up all at once;... (NG1)
   Aila forstod alt, også ting han ennå ikke hadde tenkt å si. (NG1T)
   Lit.: ... also things he yet not had thought to say

Overt non-congruent source
(3) "Since you 're going anyway, bring me a little water." (EH1T)
   "Sidan du først går, ta med litt vatn til meg." (EH1)
   Lit.: ... take with little water to me

Zero correspondence (omission in the translation)
(4) He went to fetch the scattered crutch and brought it back. (DF1)
   Han hentet den vekksparkede krykken [Ø]. (DF1T)
   Lit.: He fetched the awaykicked crutch
The corpora

<table>
<thead>
<tr>
<th>Corpora</th>
<th>English original</th>
<th>Norwegian original</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENPC+</td>
<td>1.3 mill. words</td>
<td>1.3 mill. words</td>
</tr>
<tr>
<td>BNCfiction</td>
<td>16 mill. words</td>
<td></td>
</tr>
<tr>
<td>LBKfiction</td>
<td></td>
<td>13.5 mill. words</td>
</tr>
</tbody>
</table>

- **ENPC+**: Extended version of the fiction part of the English-Norwegian Parallel Corpus (3x the size of the original ENPCfiction)
- **BNCfiction**: David Lee’s (sub-)domain W_fict_prose of the British National Corpus
- **LBKfiction**: The fiction (‘epikk’) part of Leksikografisk bokmålskorpus
Contrastive Analysis and Phraseology

Contrastive analysis (CA) is the systematic comparison of two or more languages, ...
(Johansson 2007)

Contrastive phraseology is the systematic comparison of word-combinations in two or more languages.
Approaches to phraseology

- **Phraseological approach**
  - degree of non-compositionality, or idiom status, of multiword units is one of the main concerns;
  - point of departure: often predefined sequences, falling into linguistic categories such as word-like, phrase-like, sentence-like.

- **Frequency-based approach**
  - not concerned with compositionality to the same extent and adopts a broader definition for its object of study;
  - bottom-up/inductive approach, identifying items that do not necessarily fit predefined linguistic categories;
  - in our context: unspecified items (N-grams) as starting point to identify and select patterns for research (n-gram = a contiguous sequence of n items in naturally occurring text).
Phraseology is a field bedevilled by the proliferation of terms and by conflicting uses of the same term.

What is a pattern?

- Provisional definition: a recurrent multi-word combination that functions as a semantic unit.
  - Semantics: e.g. are collocations patterns or not – *strong tea, blonde hair*
  - Morphology: e.g. are compounds patterns or not – *ivory tower, corkscrew*
  - Syntax: e.g. are phrasal verbs patterns or not – *put out*
    - (Pattern Grammar patterns (*V wh*) and collocational frames (*the x of*))
  - Discourse: e.g. are standard formulae patterns or not – *mind how you go, I don’t know*

(cf. Granger & Paquot 2008)
What is a pattern?

1. The nature of the elements involved:
   **at least one lexically specified item (form or lemma);**

2. The number of elements involved:
   **at least two;**

3. The number of times an expression must be observed:
   **greater than chance in the corpus (calculated against the number of sentences (s-units) and not orthographic words);**

4. The permissible distance between the elements involved:
   **discontinuous elements allowed, but mostly elements will be immediately adjacent, due to the starting point of continuous n-grams;**

5. The degree of lexical and syntactic flexibility of the elements involved:
   **there is usually no flexibility;**

6. The role that semantic unity and semantic non-compositionality/non-predictability play in the definition:
   (i) the pattern must be found in similar co-texts in a number of texts;
   (ii) a one-word near synonym in the source and/or target language is a good test for semantic unity;
   (iii) when two patterns are regularly translated by each other, this is a strong indication that they mean the same and have semantic unity.

(Non-compositionality is not a pre-requisite for a string of words to qualify as a pattern.)

(cf. Gries 2008)
Extended units of meaning

A language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. (Sinclair 1991: 111)

• An example: lose track
  Bella’, pleaded Marie, ‘what happened last night?’
  ‘I've lost track,’ said Bella, ‘they must have given me something ...’ (BNC)

• "The idea of a word carrying meaning on its own would be relegated to the margins of linguistic interest" (Sinclair 1996: 82).

Development of a model of extended units of meaning.
What is an extended unit of meaning?

• In addition to an **obligatory core**, the extended-unit-of-meaning model operates with four structural categories:
  – Collocation
  – Colligation
  – Semantic preference
  – Semantic prosody

• In order to discover and analyse potential extended units of meaning, it is essential to identify the (partly fixed) core.

• The **core** typically coincides with a **pattern**.
### Core: naked eye

<table>
<thead>
<tr>
<th>Location</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>130</td>
<td>H9Y 737 fast, it seemed, that I was almost invisible to the naked eye. Was it a bird? Was it a plane? No</td>
</tr>
<tr>
<td>131</td>
<td>HEH 95 get an awful lot of stars that's completely invisible to the naked eye. It picks up the very dim stars. So this this</td>
</tr>
<tr>
<td>132</td>
<td>HEH 105 've looked with binoculars and I can't see them with the naked eye. These are so faint and they only show up because of</td>
</tr>
<tr>
<td>133</td>
<td>HF1 126 I'm now seeing detail that I never saw before with the naked eye. At the moment it's mainly gall bladder and gynaecology pa</td>
</tr>
<tr>
<td>134</td>
<td>HGN 2438 that pretentious pair of shades. I look her straight in the naked eye. And look again. And yet again. Mere mortals,</td>
</tr>
<tr>
<td>135</td>
<td>HR5 1150 risen aloft, a multitude of doves — And, with the naked eye, vouchsafed visions; Watched horizontal orbs, like spotlight</td>
</tr>
<tr>
<td>136</td>
<td>HU4 1725 development of new crypts, colonies enlarge, become visible to the naked eye, then ultimately coalesce to restore the mucosal layer. This</td>
</tr>
<tr>
<td>137</td>
<td>HWT 43 adenomas and single gland adenomas, which could not be detected by naked eye, were easily recognised under the dissecting microscope (F</td>
</tr>
<tr>
<td>138</td>
<td>J0P 1110 ‘outside agitator’. Though nothing was more obvious to the naked eye than that the members of craft unions were likely to be the</td>
</tr>
<tr>
<td>139</td>
<td>JXN 508 vigorous young generation. This blood-sucking abomasal parasite, which to the naked eye is indistinguishable from H. contortus, is common in buffal</td>
</tr>
<tr>
<td>140</td>
<td>JXN 517 usually less than 7.0 mm long and difficult to see with the naked eye. Microscopic: The worms have no obvious buccal capsule.</td>
</tr>
<tr>
<td>141</td>
<td>JXN 977 bloodsuckers, have a reddish colour and can be seen by the naked eye. Superfamily STRONGYLOIDEA There are several impor</td>
</tr>
<tr>
<td>142</td>
<td>JXN 1153 inflammatory response with the formation of nodules which are visible to the naked eye. On reinfection, this response is more marked, the nodules</td>
</tr>
<tr>
<td>143</td>
<td>JXN 1444 hair-like worms 1.0-3.0 cm long which are difficult to discern with the naked eye as they are embedded in lung tissue. IDENTIFICATION T</td>
</tr>
<tr>
<td>144</td>
<td>JXN 1518 and greyish, and are not only difficult to see with the naked eye in the lung parenchyma, but are unlikely to be recovered in</td>
</tr>
<tr>
<td>145</td>
<td>JXV 1877 They're highly sophisticated infra-red beams, virtually undetectable to the naked eye. Yet another example of how well designed this whole hou</td>
</tr>
<tr>
<td>146</td>
<td>K21 2391 an ordinary card, but has differences that are invisible to the naked eye. The big banks hope it will be the answer to organised</td>
</tr>
<tr>
<td>147</td>
<td>K27 459 west, but it's too faint to be seen with the naked eye. A comet is like a huge dirty snowball, a mixture</td>
</tr>
<tr>
<td>148</td>
<td>KSW 268 A microscope. Vis., only visible, not visible to the naked eye. Look at this. Eyes undressed. What's the matter</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Collocation</td>
<td>[with / to] the naked eye</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Colligation</td>
<td>PREP + DEF. ART naked eye</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>word or phrase to do with visibility</td>
</tr>
<tr>
<td>preference</td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>difficulty (shown by words such as faint, difficult modifying visibility)</td>
</tr>
<tr>
<td>prosody</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Steps in the CA

- Produce n-gram lists on the basis of the sub-corpora of the ENPC+;
- Identify patterns;
- Select patterns for analysis;
- Identify cross-linguistic correspondences of the patterns (in original (source) and translated (target) texts) by examining concordance lines in bidirectional data;
- Perform contrastive analysis of the patterns (Milestone I);
- Investigate the co-text of the patterns in both languages on the basis of monolingual data;
- Establish – if applicable – the extended units of meaning to which the patterns belong;
- Perform contrastive analysis of the extended units of meaning (Milestone II).
Selection on the basis of n-gram lists

• The question of which n-grams to select for cross-linguistic pattern analysis relies on Gries’s criteria 3 (statistics) and 6 (semantic unity) in particular, since the extraction method automatically fulfils the other four.

• The number of potential patterns will drop after running the statistical test described in criterion 3 (statistics).
  – Follow Gries (2008: 6, n3) in multiplying the two least frequent words of the gram divided by the total number of s-units in the corpus at hand.

• Criterion 6 (semantic unity) cannot be completely fulfilled unless some analysis is performed.

• By way of illustration, the 4-gram *out of the ordinary* is selected.
Selection of pattern to study

- The sequence *out of the ordinary* in English original texts satisfies all of the six defining criteria;

1. Nature of the elements: *lexically specified*
2. Number of elements: *four*
3. Frequency: *chance frequency (3.6) < actual frequency (17)*
4. Distance bw elements: *adjacent*
5. Lexical/syntactic flex.: *no flexibility*
6. Semantic unity: *one-word synonym: unusual (OxfordDictionaries); one-word correspondence in another language, e.g. Norwegian *uwanlig 'unusual', Czech neobvyklý 'unusual'*
   a comparable sequence of words in another language, e.g. Norwegian *utenom det vanlige 'outside (different from) the ordinary'*
One Czech example

Did you hear anything -- any sounds, or anything out of the ordinary?

Slyšeli jste neco? Nejaké zvuky nebo neco neobvyklého?

(OPUS - OpenSubtitles Corpus)
Nothing out of the ordinary. [WB1]

...mell of joss-sticks, but if she behaved as if nothing was out of the ordinary, all would be fine. [JB1]

Jeg fortalte at han antagelig hadde gurvinnene for og at det korn til å lukte røkelse, men at alt ville gå bra om hun lot som ingenting.

Nothing out of the ordinary... and nothing we would n't ask a while man in the same case. [MiWa1E]

Ikke noe uvanlig... og ingenting vi ikke ville spare en hvit mann om under samme omstendigheter..."

Well, she did behave as if nothing was out of the ordinary, and I began to suspect that Oliver was a little displeas... [JB1]

Vel, hun lot virkelig som ingenting, og jeg begynte å ane at Oliver ble litt skuffet.

"So nothing out of the ordinary happened on this trip you and Mark Hardcastle made to London... [PeRo2E]

"Det skjedde altså ingenting uvanlig på denne turen De og Mark Hardcastle tok til London?" sa Banks.

...were shattered is unrecognisable as a place where anything out of the ordinary happened. [StGa1E]

Men når mennesker dør, blir de fjernet, brukt til sykehus og gravlunder, og før kroppene er leget eller kalde er stedet der de ble ødelagt like ubetydelig å se på som ethvert annet sted der intet spesielt er skjedd.

"We've checked all the footage, and there's nothing out of the ordinary," he said. [PeRo2E]

"Vi har sjekket alle optakene, og det er ingenting utover det vanlige," sa han.

...could see that the keys were gone, and there was nothing out of the ordinary in the interior on a cursory examination in poor light con... [PeRo2E]

Hun tørt av regnet og kikket inn gjennom vinduet. Næklene var borte, og det var ingenting uvanlig å se innenfor ved en overfladisk undersøkelse i det dårlige lyset.

Nothing out of the ordinary on it so far. [PeRo2E]

Ikke noe ut over det vanlige på den, så vidt vi kan se.
Cross-linguistic correspondences

<table>
<thead>
<tr>
<th>Norwegian correspondences of <em>out of the ordinary</em></th>
<th>Norwegian translation</th>
<th>Norwegian source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Uvanlig</em> ‘unusual’</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td><em>Utover det vanlige</em> ‘beyond (in addition to) the ordinary’</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td><em>(ikke) normal</em> ‘(not) normal’</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td><em>Utenom det vanlige</em> ‘outside (different from) the ordinary’</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><em>Unormal</em> ‘un-normal’</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>Påfallende</em> ‘remarkable’</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td><em>Spesiell</em> ‘special’</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>other</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>4</td>
</tr>
</tbody>
</table>
(1) ... om en bok han ganske enkelt måtte lese, en bok helt utenom det vanlige, naturligvis, om aksjer,... [OEL1]
... about a book he simply must read, a book quite out of the ordinary, of course, about stocks and shares, ... [OEL1T]

(2) Well, she did behave as if nothing was out of the ordinary, and I began to suspect that ... [JB1]
Vel, hun lot virkelig som ingenting, og jeg begynte å ane at ... [JB1T]
Lit.: Well, she pretended really as nothing (=she acted as if nothing had happened), and I began to suspect that...

(3) Nothing out of the ordinary ... and nothing we wouldn’t ask a white man in the same ... [MiWa1E]
Ikke noe uvanlig ... og ingenting vi ikke ville spørre en hvit mann om under samme ... [MiWa1TN]
Lit.: Not something unusual ...

(4) Keep a lookout for anything out of the ordinary. [PeRo2E]
Være på vakt overfor alt som ikke virker normalt. [PeRo2TN]
Lit.: Be on guard opposite everything that not seems normal
Top three Norwegian correspondences

- In the Norwegian original texts of the ENPC+
  - *utover det vanlige* is nonexistent
  - *utenom det vanlige* is relatively infrequent (4 occ.)
  - *uovanlig* is relatively frequent (67 occ.)

- The mutual correspondence is low:
  - 7% for *uovanlig* and *out of the ordinary*
  - 15% for *utenom det vanlige* and *out of the ordinary*
Mutual Correspondence (MC)

The frequency with which different (grammatical, semantic and lexical) expressions are translated into each other. (Altenberg 1999)

In the case of *out of the ordinary* and *uvanlig*:

\[
\frac{(6+0) \times 100}{17 + 67} = 7.14\% 
\]

\(A_t\) and \(B_t\) = The number of times the compared items (A and B) are translated by each other

\(A_s\) and \(B_s\) = The total number of occurrences of the compared items in the source/original texts
A low MC may indicate that there is a phraseological gap between the two languages, as no specific pattern or single word stands out as the correspondence of out of the ordinary.
Extended units of meaning in contrast

- Sinclair’s model of extended units of meaning is used to test the equivalence of these cross-linguistic correspondences in terms of similarity of meaning in context.

- Focus on the congruent correspondences *out of the ordinary* and *utenom det vanlige*
<table>
<thead>
<tr>
<th>N–3</th>
<th>N–2</th>
<th>N–1</th>
<th>Node</th>
<th>N+1</th>
<th>N+2</th>
<th>Translation (of node)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lookout</td>
<td>for</td>
<td>anything</td>
<td>out of the ordinary</td>
<td>.</td>
<td></td>
<td>normalt</td>
</tr>
<tr>
<td>n’t</td>
<td>see</td>
<td>anything</td>
<td>out of the ordinary</td>
<td>.</td>
<td></td>
<td>utover det vanlige</td>
</tr>
<tr>
<td>place</td>
<td>where</td>
<td>anything</td>
<td>out of the ordinary</td>
<td>happened</td>
<td>.</td>
<td>spesielt</td>
</tr>
<tr>
<td>to</td>
<td>know</td>
<td>anything</td>
<td>out of the ordinary</td>
<td>that happened</td>
<td>.</td>
<td>uvanlig</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nothing</td>
<td>out of the ordinary</td>
<td>...</td>
<td>and</td>
<td>uvanlig</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nothing</td>
<td>out of the ordinary</td>
<td>on</td>
<td>it</td>
<td>ut over det vanlige</td>
</tr>
<tr>
<td></td>
<td></td>
<td>So</td>
<td>out of the ordinary</td>
<td>happened</td>
<td>on</td>
<td>uvanlig</td>
</tr>
<tr>
<td></td>
<td></td>
<td>computer</td>
<td>showed nothing</td>
<td>out of the ordinary</td>
<td>on</td>
<td>a</td>
</tr>
</tbody>
</table>
(5) ‘Now, this is something really out of the ordinary.’ [BNC: AEB 3153]
Extended context of *utenom det vanlige* and *out of the ordinary*

- Based on data from LBKfiction/ENPC+ and BNCfiction/ENPC+ (22 occ. / 58 occ.)
(6) Vi vil ikke at forbipasserende skal tro at det har skjedd noe utenom det vanlige. [LBK: BoEl01.350]
Lit.: We want not that passers-by shall believe that there has happened something out of the ordinary.

(7) But when people die they’re removed, taken to hospitals and graveyards, and before the bodies are healed or cold the spot where they were shattered is unrecognisable as a place where anything out of the ordinary happened. [StGa1E]
• Both patterns are often found in combination with non-assertive pronouns;
• This tendency is much more prominent in the case of out of the ordinary, pointing to a difference not only with regard to co-occurrence but also with regard to semantic prosody;
• The Norwegian pattern is more commonly found in a positive context and can thus be claimed to be a semantic equivalent of the extended unit of meaning with out of the ordinary where it has a positive semantic prosody;
• This is corroborated by translational evidence from the ENPC+, where the only instance of utenom det vanlige corresponding to out of the ordinary is the one where something is its left collocate:
(8) Perhaps if there was something out of the ordinary they made it disappear? [PeRo2E]
Hvis det var noe utenom det vanlige der, så kanskje de fikk det til å forsvinne? [PeRo2TN]
Degree of similarity between (nothing/anything) out of the ordinary and *utenom det vanlige*
Degree of similarity between (something) *out of the ordinary* and *utenom det vanlige*
Out of the ordinary vs. uvanlig

- A similar analysis for out of the ordinary and its most frequent Norwegian correspondence, viz. uvanlig, gives a different picture;

- There is no evidence to support the claim that uvanlig is the node of a similar extended unit of meaning;

- Uvanlig may be too syntactically versatile and not stable enough in one particular environment for this to be the case;

- Due to this flexible nature of uvanlig, it may work as a non-congruent correspondences of out of the ordinary in many cases.
Conclusion: Case study

- The method of contrastive pattern analysis can be successfully applied even on relatively infrequent patterns;
- The cross-linguistic analysis of *out of the ordinary* uncovered several Norwegian correspondences, many of which are not found in English-Norwegian bilingual dictionaries;
- One of the correspondences – *utenom det vanlige* – was established as a pattern and its degree of similarity with *out of the ordinary* was measured;
- *Out of the ordinary* shows some overlap with a Norwegian pattern and a Norwegian one-word item, but only the extended unit of meaning with a positive prosody has a close match in Norwegian, namely (*noe*) *utenom det vanlige* = *(something)* *out of the ordinary*.
- This is a factor contributing to the low MC of the two units, since the positive version was not attested to the same extent in the ENPC+. 
Concluding remarks

• How to go from “raw” corpus data to meaning (in context) in a contrastive perspective;
• The bidirectional model has proved to be useful when applying a bootstrapping method, extracting cross-linguistic n-grams for the study of patterns and extended units of meaning;
• Important new findings emerge as a result of combining the fields of contrastive linguistics and phraseology;
• The method of extracting, identifying and comparing word combinations in a systematic way contributes to an understanding of how patterns work cross-linguistically.
• “Corpus linguistics in a new key”; the corpora have been exploited “for language comparison and for throwing special features of the languages compared into relief, including preferred ways of expressing similar things” (Johansson 2012: 64).
We are better equipped to “see the characteristics of each language more clearly, and the comparison can contribute to a better description of each individual language. This type of CA has sometimes been called analytic comparison or linguistic characterology (Mathesius 1975)” (Johansson 2008: 10).
THANK YOU!
References


Corpora

British National Corpus (BNC). http://www.natcorp.ox.ac.uk/

English-Norwegian Parallel Corpus (ENPC). http://www.hf.uio.no/ilos/english/services/omc/enpc/

Leksikografisk bokmålskorpus (LBK). http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/