Metaphorical and Literal Language in Translation

- Metaphors are a hard nut to crack: even more so beyond monolingual settings.
- Contrast between metaphorical and literal language in translation and its potential effect on translatability and variability in language production and generation.
- Translation modes.

**VOLIMET**, a corpus of 2,916 parallel sentences containing gold standard en2de and en2fr alignments of metaphorical verb-object pairs and their literal paraphrases, e.g., tackle/address question.

- The parallel nature of our corpus enables us to:
  - Explore monolingual patterns for metaphorical vs. literal uses in English.
  - Investigate different aspects of cross-lingual translations into German and French and the extent to which metaphoricity and literalness in the source language are transferred to the target languages.

Creating VOLIMET

1. 47 synonym VO met vs. lit English pairs
2. 295 VO pairs extension e.g., fill vs. cancel proposals
   - en2fr + en2de EUROPAL CORPUS
   - automatic alignments

HUMAN ANNOTATIONS:
- linking alignments
- specific guidelines

Monolingual Analyses

<table>
<thead>
<tr>
<th>Met. VO pairs</th>
<th>Lit. VO pairs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>instances</td>
<td>instances</td>
<td></td>
</tr>
<tr>
<td>730 (12.9%)</td>
<td>961 (10.9%)</td>
<td>1,691</td>
</tr>
<tr>
<td>VO pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58 (27)</td>
<td>88 (32)</td>
<td>144</td>
</tr>
<tr>
<td>Inflected VOs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>135 (23.3)</td>
<td>203 (23.3)</td>
<td>338</td>
</tr>
<tr>
<td>Avg. sent. len</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>–</td>
<td>34.16</td>
<td>–</td>
</tr>
</tbody>
</table>

Statistics on extracted monolingual English data:
- number of instances containing metaphorical and literal VOs:
- number of extracted VOs:
- number of extracted VOs

Imbalances in the frequencies of VOs, e.g., 2 instances of shape outcome, but 169 instances of tackle problem.

Crosslingual Analyses

<table>
<thead>
<tr>
<th>Met. VO pairs</th>
<th>Lit. VO pairs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># parallel sentences</td>
<td>719 (12.48)</td>
<td>982 (11.55)</td>
</tr>
<tr>
<td># VO pairs</td>
<td>58 (27)</td>
<td>88 (32)</td>
</tr>
<tr>
<td># inflected VOs</td>
<td>135 (23.3)</td>
<td>203 (23.3)</td>
</tr>
<tr>
<td>Avg. sent. len</td>
<td>50.88</td>
<td>34.16</td>
</tr>
</tbody>
</table>

- The obtained data is overall balanced for en2de. However, the en2fr parallel dataset containing source metaphoric VO pairs is twice as large as the one containing source literal VOs.
- Average of 11 parallel sentences per VO. But variations across VOs. For instance, for the metaphorical en2fr dataset, we obtained only one parallel sentence containing the source VO break contract but 102 parallel sentences containing the source VO find way.
- VARIABILITY IN TRANSLATION
  - Avg. 280 unique translations across language pairs, we find only 19 en2fr and 27 en2de translations that are translations of several metaphorical VOs, e.g., tackle challenge and tackle issue: répondre à une question
  - Variation of unique translations across VOs; e.g., 302 en2de parallel sentences containing the VO address problem: 79 different (unique) translations. The number of instances per VO is highly correlated with the number of translations (avg. Spearman’s correlation ρ=0.99 for en2de and ρ=0.88 for en2fr), i.e., the more a VO appears in natural language, the more (unique) translations are produced.
- SIZE AND FREQUENCY
  - The number of sentences containing metaphorical and literal VOs is significantly higher in the source language than in the target language.
- SYNTACTIC VARIATION
  - The number of unique translations per VO is highly correlated with the number of translations (Spearman’s correlation ρ=0.99 for en2de and ρ=0.88 for en2fr), i.e., the more a VO appears in natural language, the more (unique) translations are produced.

Variations of VO pairs: metaphorical (blue) vs. literal (orange).

Out of the 58 metaphorical and 88 literal retrieved VOs, 31 of them are actual paraphrased pairs whose frequencies can be compared.

Takeaways

- **VOLIMET**, the first parallel corpus of English–German and English–French paraphrased metaphorical and literal verb-object pairs.
- Gold standard alignments of source VO pairs to all their corresponding translations.
- Mono-lingually, there exists a clear preference for either the metaphorical or the literal variant for some VO pairs.
- Cross-lingually, our findings revealed substantial variability in translations, i.e., one-to-many mappings between source VO pairs and their target translations.
- French translations show equal use of metaphorical and literal language, while German tends to favor literal translations by a large margin.

- Crosslingual Analyses

<table>
<thead>
<tr>
<th>Met. VO pairs</th>
<th>Lit. VO pairs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># parallel sentences</td>
<td>719 (12.48)</td>
<td>982 (11.55)</td>
</tr>
<tr>
<td># VO pairs</td>
<td>58 (27)</td>
<td>88 (32)</td>
</tr>
<tr>
<td># inflected VOs</td>
<td>135 (23.3)</td>
<td>203 (23.3)</td>
</tr>
<tr>
<td>Avg. sent. len</td>
<td>50.88</td>
<td>34.16</td>
</tr>
</tbody>
</table>

- The obtained data is overall balanced for en2de. However, the en2fr parallel dataset containing source metaphoric VO pairs is twice as large as the one containing source literal VOs.
- Average of 11 parallel sentences per VO. But variations across VOs. For instance, for the metaphorical en2fr dataset, we obtained only one parallel sentence containing the source VO break contract but 102 parallel sentences containing the source VO find way.
- VARIABILITY IN TRANSLATION
  - Avg. 280 unique translations across language pairs, we find only 19 en2fr and 27 en2de translations that are translations of several metaphorical VOs, e.g., tackle challenge and tackle issue: répondre à une question
  - Variation of unique translations across VOs; e.g., 302 en2de parallel sentences containing the VO address problem: 79 different (unique) translations. The number of instances per VO is highly correlated with the number of translations (avg. Spearman’s correlation ρ=0.99 for en2de and ρ=0.88 for en2fr), i.e., the more a VO appears in natural language, the more (unique) translations are produced.
- SIZE AND FREQUENCY
  - The number of sentences containing metaphorical and literal VOs is significantly higher in the source language than in the target language.
- SYNTACTIC VARIATION
  - The number of unique translations per VO is highly correlated with the number of translations (Spearman’s correlation ρ=0.99 for en2de and ρ=0.88 for en2fr), i.e., the more a VO appears in natural language, the more (unique) translations are produced.

Variations of VO pairs: metaphorical (blue) vs. literal (orange).

Out of the 58 metaphorical and 88 literal retrieved VOs, 31 of them are actual paraphrased pairs whose frequencies can be compared.

Takeaways

- **VOLIMET**, the first parallel corpus of English–German and English–French paraphrased metaphorical and literal verb-object pairs.
- Gold standard alignments of source VO pairs to all their corresponding translations.
- Mono-lingually, there exists a clear preference for either the metaphorical or the literal variant for some VO pairs.
- Cross-lingually, our findings revealed substantial variability in translations, i.e., one-to-many mappings between source VO pairs and their target translations.
- French translations show equal use of metaphorical and literal language, while German tends to favor literal translations by a large margin.