Generalization in Multilingual Semantic Parsing

Workshop on Ten Years of BabelNet and Multilingual Neurosymbolic Natural Language Understanding
5 July 2022

Daniel Hershcovitch
# Why Meaning Representation for NLP?

## Performance
- Inductive bias
- Access to structured data
- Reasoning ability

## Understanding
- Interpretability
- Theoretical analysis
- Fine-grained control

## Generalization
- Languages
- Domains
- Tasks
Graph-based Meaning Representations

<table>
<thead>
<tr>
<th>Type of Information</th>
<th>DM</th>
<th>PSD</th>
<th>EDS</th>
<th>DMRS</th>
<th>UCCA</th>
<th>AMR</th>
<th>PMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicates–Arguments</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Sense Differentiation</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

Graph-Based Meaning Representations: Design and Processing (Koller et al., ACL 2019)

Multitask Parsing Across Semantic Representations (Hershcovich et al., ACL 2018)
Universal Conceptual Cognitive Annotation (UCCA)

Design principles

- Cross-linguistic portability and stability
- Accessibility to non-expert annotators
- Modularity of semantic components

Corpora

- English, German, French, Russian, Hebrew & now also Turkish

Universal Conceptual Cognitive Annotation (UCCA) (Abend & Rappoport, ACL 2013)
Turkish Universal Conceptual Cognitive Annotation (Bölücü & Can, LREC 2022)
After graduation, John moved to Paris.

ג'ו' עברה לפרו אט' אט'شت רבי'راه', הלימודים

F P D L H A P A
UCCA Parsing

Successful cross-lingual transfer

Meaning Representation Parsing

Similar results across languages

MRP 2020: The Second Shared Task on Cross-Framework and Cross-Lingual Meaning Representation Parsing (Oepen et al., CoNLL 2020)
Cross-lingual Generalization

Dominant approaches:

Multilingual models (mBERT, XLM-R etc.)

Annotation projection (manual/automatic translation)
**Compositional Generalization**

"THE ABILITY TO SYSTEMATICALLY GENERALIZE TO COMPOSED TEST EXAMPLES OF A CERTAIN DISTRIBUTION AFTER BEING EXPOSED TO THE NECESSARY COMPONENTS DURING TRAINING ON A DIFFERENT DISTRIBUTION"

<table>
<thead>
<tr>
<th>Train set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who directed Inception?</td>
<td>Did Greta Gerwig direct Goldfinger?</td>
</tr>
<tr>
<td>Did Greta Gerwig produce Goldfinger?</td>
<td>Who produced Inception?</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
CFQ (Compositional Freebase Questions)

**English**

Did a film editor's parent direct M0?

**SPARQL**

```sparql
SELECT count(*) WHERE {
  ?x0 ns:film.director.film [M0] .
  ?x0 ns:people.person.children ?x1 .
  ?x1 a ns:film.editor
}
```

Measuring compositional generalization: A comprehensive method on realistic data (Keysers et al., ICLR 2020)
CFQ (Compositional Freebase Questions)

Measuring compositional generalization: A comprehensive method on realistic data (Keysers et al., ICLR 2020)
Multilingual Compositional Wikidata Questions

Ruixiang Cui
Rahul Aralikatte
Heather Lent

CFQ
(239K instances)

Freebase

Property mapping
Entity substitution

English

Automatic & manual translation

MCWQ
(124K instances)

Wikidata

Hebrew, Kannada, Chinese & English

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
### Multilingual Compositional Wikidata Questions

<table>
<thead>
<tr>
<th>Lang.</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>En</td>
<td>Did Lohengrin’s male actor marry Margarete Joswig</td>
</tr>
<tr>
<td>He</td>
<td>האם של לוהנגרן הגר护身符ו של גבר של הגר在线咨询</td>
</tr>
<tr>
<td>Kn</td>
<td>సమాన పారముఖం హిందు సాధన సాధన బాధాంచన మాధార్స్</td>
</tr>
<tr>
<td>Zh</td>
<td>Lohengrin 的 男 演员 嫁给了 Margarete Joswig 吗</td>
</tr>
</tbody>
</table>

**SPARQL Query:**

```sparql
ASK WHERE { ?x0 wdt:P453 wd:Q50807639 . ?x0
FILTER ( ?x0 != wd:Q1560129 )}
```

[Compositional Generalization in Multilingual Semantic Parsing over Wikidata](Cui et al., to appear in TACL)
Multilingual Compositional Generalization

<table>
<thead>
<tr>
<th>Within-language</th>
<th>Random split</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>En</td>
</tr>
<tr>
<td>Exact Match (%)</td>
<td></td>
</tr>
<tr>
<td>LSTM+Attention</td>
<td>96.6</td>
</tr>
<tr>
<td>E. Transformer</td>
<td>99</td>
</tr>
<tr>
<td>mBERT</td>
<td>98.7</td>
</tr>
<tr>
<td>T5-base+RIR</td>
<td>98.5</td>
</tr>
<tr>
<td>mT5-small+RIR</td>
<td>98.6</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td><strong>99.1</strong></td>
</tr>
</tbody>
</table>

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
## Multilingual Compositional Generalization

<table>
<thead>
<tr>
<th>Within-language</th>
<th>( \text{MCD}_{\text{mean}} )</th>
<th>Random split</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Exact Match (%)</strong></td>
<td>En</td>
<td>He</td>
</tr>
<tr>
<td>LSTM+Attention</td>
<td>19.4</td>
<td>15.5</td>
</tr>
<tr>
<td>E. Transformer</td>
<td>29.3</td>
<td>18.9</td>
</tr>
<tr>
<td>mBERT</td>
<td>26.6</td>
<td>22.7</td>
</tr>
<tr>
<td>T5-base+RIR</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>mT5-small+RIR</td>
<td><strong>38.3</strong></td>
<td>29.3</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td>33.8</td>
<td><strong>33.2</strong></td>
</tr>
</tbody>
</table>

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
Multilingual Compositional Generalization

<table>
<thead>
<tr>
<th>Within-language</th>
<th>MCD\textsubscript{mean}</th>
<th>Random split</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>En</td>
<td>He</td>
</tr>
<tr>
<td>Exact Match (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSTM+Attention</td>
<td>19.4</td>
<td>15.5</td>
</tr>
<tr>
<td>E. Transformer</td>
<td>29.3</td>
<td>18.9</td>
</tr>
<tr>
<td>mBERT</td>
<td>26.6</td>
<td>22.7</td>
</tr>
<tr>
<td>T5-base+RIR</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>mT5-small+RIR</td>
<td>38.3</td>
<td>29.3</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td>33.8</td>
<td>33.2</td>
</tr>
</tbody>
</table>

Cross-lingual (from English)

<table>
<thead>
<tr>
<th></th>
<th>MCD\textsubscript{mean}</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>En</td>
<td>He</td>
</tr>
<tr>
<td>mT5-small+RIR</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td>0.4</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
# Multilingual Compositional Generalization

## Within-language

<table>
<thead>
<tr>
<th>Exact Match (%)</th>
<th>MCD\textsubscript{mean}</th>
<th>Random split</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>En</td>
<td>He</td>
</tr>
<tr>
<td>LSTM+Attention</td>
<td>19.4</td>
<td>15.5</td>
</tr>
<tr>
<td>E. Transformer</td>
<td>29.3</td>
<td>18.9</td>
</tr>
<tr>
<td>mBERT</td>
<td>26.6</td>
<td>22.7</td>
</tr>
<tr>
<td>T5-base+RIR</td>
<td>28.1</td>
<td></td>
</tr>
<tr>
<td>mT5-small+RIR</td>
<td>38.3</td>
<td>29.3</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td>33.8</td>
<td>33.2</td>
</tr>
</tbody>
</table>

## Cross-lingual (from English)

### SPARQL BLEU

<table>
<thead>
<tr>
<th></th>
<th>En</th>
<th>He</th>
<th>Kn</th>
<th>Zh</th>
<th>En</th>
<th>He</th>
<th>Kn</th>
<th>Zh</th>
</tr>
</thead>
<tbody>
<tr>
<td>mT5-small+RIR</td>
<td>87.5</td>
<td>53.8</td>
<td>53.2</td>
<td>59</td>
<td>99.9</td>
<td>60.4</td>
<td>59.9</td>
<td>63.8</td>
</tr>
<tr>
<td>mT5-base+RIR</td>
<td>86.4</td>
<td>46.4</td>
<td>46</td>
<td>52.7</td>
<td>99.9</td>
<td>63.2</td>
<td>63.5</td>
<td>70.6</td>
</tr>
</tbody>
</table>

### Exact Match (%)

<table>
<thead>
<tr>
<th></th>
<th>mT5-small+RIR</th>
<th>mT5-base+RIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.2</td>
<td>0.3</td>
</tr>
</tbody>
</table>

---

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
Error Analysis

Cross-lingual prediction errors

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
Error Analysis

Question: Was M0 written by and directed by M1, M2, and M3

Gold:

```
```

Inferred:

```
ASK WHERE { M0 wdt:P57 M1 . M1 wdt:P57 M2 . M0 wdt:P58 M3 }
```

Incorrect predicate-argument structure

Compositional Generalization in Multilingual Semantic Parsing over Wikidata (Cui et al., to appear in TACL)
Can Multilingual Meaning Representation Help?
Summary

- ✔️ Multilingual compositional generalization benchmark

- ❗️ Similar within-language generalization across languages

- ✖️ Zero-shot cross-lingual generalization fails
Limitations & Future Work

• Synthetic & unnatural (inherited from CFQ)
  Paraphrasing & expansion

• Mostly automatic translation
  Improvement with RBMT

• No cultural adaptation
  Native data collection

Challenges and Strategies in Cross-Cultural NLP (Hershcovich et al., ACL 2022)
Thanks!
danielhers.github.io
@daniel_hers
dh@di.ku.dk
Semantics for Analysis

**Context:** A piece of paper was later found on which he had written his last statements in **two** languages, Latin and German. **Only one** statement was in Latin and the rest in German.

**Question:** In what language were **most** statements written? **Answer:** German  **Predicted Answer:** Latin and German

**Premise:** Más de tres personas resultaron heridas en un accidente de dos vehículos el lunes por la noche. (translation: **More than three** people were injured in a **two-vehicle** crash Monday evening.)

**Hypothesis:** Había 4 personas involucradas. (translation: **There were 4** people involved. **Label:** Neutral  **Predicted Label:** Entailment

*[Generalized Quantifiers as a Source of Error in Multilingual NLU Benchmarks](https://example.com)* (Cui et al., NAACL 2022)
Semantics for Interpretability

**NL Questions over Information Sources**

- "What color are a majority of the objects?" (Image)
- "From what yard line did Shayne Graham kick two field goals?" (Passage)
- "Return the keywords which have been contained by more than 100 ACL papers." (Database)

**Question Decomposition**

1. objects
2. colors of #1
3. number of #1 for each #2
4. #2 where #3 is highest
5. #3 where #4 is highest

1. Shayne Graham
2. field goals of #1
3. yards of #2
4. number of #2 for each #3
5. #3 where #4 is more than 100

**Decomposition Graph**

- superlative[max]
- group[count]
- project[colors]
- select[objects]

- comparative[=,two]
- group[count]
- project[yards]
- select[field goals]

- comparative[>,100]
- group[count]
- project[keywords]
- filter[ACL]
- select[papers]

**Break It Down: A Question Understanding Benchmark** (Wolfson et al., TACL 2020)
UCCA
Implicit Relations in UCCA

Refining Implicit Argument Annotation for UCCA (Cui & Hershcovich, DMR 2020)
Great Service! Fine-grained Parsing of Implicit Arguments (Cui & Hershcovich, IWPT 2021)
IBM happened to choose a company with a crucial vulnerability, despite vetting.
Graph-Based Meaning Representations: Design and Processing (Koller et al., ACL 2019)
Parsing Progress

AMR 2017 Smatch F1 [Link](https://paperswithcode.com/sota/amr-parsing-on-ldc2017t10)
Designing a Uniform Meaning Representation for Natural Language Processing (Van Gysel et al., KI - Künstliche Intelligenz 2021)
UMR-Writer: A Web Application for Annotating Uniform Meaning Representations (Zhao et al., EMNLP 2021)