**INTRODUCTION**

Language grounding aims at linking the symbolic representation of language (e.g., words) into the rich perceptual knowledge of the physical world.

**ARCHITECTURE**

- Trained on COCO dataset with parallel multilingual captions.
- The alignment M is trained on a limited number of words (those that occur in the captions), then applied to all the textual vectors to generate “zero-shot” grounded embeddings.

**QUANTITATIVE RESULTS**

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>WSim</th>
<th>MEN</th>
<th>RW</th>
<th>MTurk</th>
<th>SimVerb</th>
<th>SimLex</th>
</tr>
</thead>
<tbody>
<tr>
<td>English/Arabic</td>
<td>3.9%</td>
<td>5.4%</td>
<td>4.6%</td>
<td>6.8%</td>
<td>1.8%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Arabic/English</td>
<td>6.9%</td>
<td>6.9%</td>
<td>9.3%</td>
<td>7.0%</td>
<td>9.3%</td>
<td>7.0%</td>
</tr>
<tr>
<td>German/English</td>
<td>9.6%</td>
<td>10.4%</td>
<td>6.0%</td>
<td>6.3%</td>
<td>9.6%</td>
<td>10.4%</td>
</tr>
</tbody>
</table>

 Improvement in Pearson correlation (%) of the grounded embeddings compared to their textual counterparts on unsupervised semantic similarity benchmarks. The table shows the result of grounding a single language (left) vs. the addition of a second language (right).

**QUALITATIVE RESULTS**

Out of the top 10 nearest neighbors for each query word, only the differing neighbors between the textual embeddings and the grounded embeddings are reported.

**CONCLUSION**

- Grounding improves embeddings in all three languages.
- Similar languages benefit from each other, but differing languages seem to conflict on some aspects.
- A more advanced architecture is needed to link the three languages.
- Analysing inter-lingual grounding in fine granularity is to be investigated.