**Problem: What is skip-gram really doing?**
1. Skip-gram (Mikolov 2013) is just exponential family principal components analysis (EPCA) formulated in Collins (2001)

\[
\sum_j \sum_i X_{ij} \log p(\text{context } i \mid \text{word } j) = \sum_j \sum_i X_{ij} \log \frac{\exp (c_i \cdot w_j)}{\sum_{j'} \exp (c_{i'} \cdot w_j)}
\]

2. EPCA is a form of matrix factorization, so **Higher-Order Skip-Gram** uses tensor factorization to include more information to create word embeddings.

**Multi-Lingual Word Similarity for Morphology Tensor**

Word similarity results comparing the compositional morphology tensor with the standard skip-gram model.

**Correlation with UD Part-of-Speech Annotation in 40 Languages for Positional Tensor**

<table>
<thead>
<tr>
<th>Language</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
</tr>
</tbody>
</table>

**Higher-Order Skip-Gram: Tensor Factorization Extension**

\[
||X - C \otimes W \otimes R||_F^2 = \sum_{ijk} (x_{ijk} - 1 \cdot (c_i \otimes w_j \otimes r_k))^2
\]

**Conclusion**

1. Skip-Gram can be extended to **tensor factorization**
2. Tensor-factorization enables the use of **more linguistic structure** when creating word embeddings
3. Higher-Order Skip-Gram achieved **better word embeddings** as evaluated on 40 languages