Overview

Morphological compounding is prevalent and systematic across many of the world's languages. By examining how a particular concept is compounded across languages, can we predict how it is compounded in a specific language?

By using only freely-available dictionaries, we can gather plausible compounds and construct compounding recipes. We use these recipes to discover additional compounds and identify common compounding glue characters.

Compounding Mechanisms

We consider various mechanisms to combine two components into a single compound word:

Simple concatenation: A + B
- fi: rakennustyö "construction" = construction + work
- nl: ziekenhuis "hospital" = sick + house

Glue Character: A + glue + B
- da: folkeformening "reform" = people + vote
- nb: informasjonsteknologi "information technology"

Drop Left: A[-1] + B
- de: meuchelmörder "assassin" = assassinate + killer
- es: cantautor "singer-songwriter" = singing + author

Experiments

We evaluate the effectiveness of our compound recipes in handling unknown words (held-out from the dictionary) in two directions:

- e2f: given a concept, can we generate the foreign compound word?
- f2e: given a foreign compound word, can we analyze it and identify the concept?

Sample Results

<table>
<thead>
<tr>
<th>Lang</th>
<th>English Concept</th>
<th>Foreign</th>
<th>Foreign Analysis</th>
<th>Literal Translation</th>
<th>Top Hyp</th>
<th>2nd Hyp</th>
<th>Rank</th>
<th>f2e</th>
</tr>
</thead>
<tbody>
<tr>
<td>eo</td>
<td>yolk</td>
<td>ovoflavo</td>
<td>ovo + flavo</td>
<td>egg yellow</td>
<td>yolk</td>
<td>egg yolk</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>wyi</td>
<td>skull</td>
<td>galk</td>
<td>galk + 'g' + gawang</td>
<td>bone head</td>
<td>skull</td>
<td>cranium</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>mns</td>
<td>fifty</td>
<td>atlov</td>
<td>at + lov</td>
<td>five ten</td>
<td>fifteen</td>
<td>fifty</td>
<td>2</td>
<td>Y</td>
</tr>
<tr>
<td>nb</td>
<td>continent</td>
<td>fastland</td>
<td>fast + land</td>
<td>firm country</td>
<td>mainland</td>
<td>continent</td>
<td>2</td>
<td>Y</td>
</tr>
</tbody>
</table>

Code and data at github.com/wswu/worcomal