Understanding out-prefixation: merging qualitative and distributional analyses

Martin Schäfer and Sven Kotowski

Universität Tübingen and Heinrich-Heine-Universität Düsseldorf

19th International Morphology Meeting, Wien, 07.02.2020
Introduction

(1) a. The German Me 262 jets could **outfly** the Mustangs by 100 MPH in level flight. (iWeb)
b. The point of A-Wing interceptors wasn’t technically combat superiority. There are better interceptors [...] In terms of maneuverability, there wasn’t much that could **outfly** one. (iWeb)

(2) The flow is a slow drawl to fit with the song and I like that you’re not **outrapping** the beat. (iWeb)
eventuality

\[ \begin{array}{c}
\text{ACTOR} \lor \text{UNDERGOER} \\
\text{M-DIM} \begin{bmatrix}
\text{scale type} \land \text{dimension} \\
\text{SCALAR ATTRIBUTE}_i \alpha
\end{bmatrix}
\end{array} \]
What about these scales?

- Ahn 2018, Tolskaya 2014: the lexical predicates help to identify the scale of comparison
- Our talk: testing this empirically
  1. Part 1:
     - Study 1: Verbnet classes and scalar dimensions: higher level generalizations
     - Study 2: Verbnet classes and scalar dimensions: lemma-level reflections
  2. Part 2: Distributional similarity within Verbnet classes and between bases and derivations
Classes of out-prefixed verbs and their scalar dimensions

- iWeb (Davies 2018)
  web-based corpus; 14 billion words/ 95,000 websites; different varieties

- Verb-classification: VerbNet (see Kipper et al. 2008; Levin 1993) Database of 8537 English verbs
  - 273 main classes
  - based on syntactic alternation frames/role information on possible arguments
  - argued to show core semantic similarity
Study 1

- all out-prefixed lemmas with morphological bases from seven VerbNet-classes: PERFORMANCE, RUN, EXIST, SPRAY, SOUND-EMISSION, CARRY, and HIT

- 450 verbs checked, 104 out-verb-types

<table>
<thead>
<tr>
<th>run-verbs</th>
<th>performance-verbs</th>
<th>exist-verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>super-ordinate description</td>
<td>Verbs of Motion</td>
<td>Verbs of Creation and Transformation</td>
</tr>
<tr>
<td>non-directional manner of movement</td>
<td>performances can occur as effected objects</td>
<td>existence of an entity at some location</td>
</tr>
<tr>
<td><em>crawl, creep, run, jump</em> etc.</td>
<td><em>chant, play, dance, sing</em> etc.</td>
<td><em>dwell, exist, live, remain</em> etc.</td>
</tr>
</tbody>
</table>
Study 1

Do VerbNet-classes allow for higher-level generalizations wrt dimension-based comparisons?

- combinations of *out*-lemmas with concrete information on scalar dimensions
- combinations of verb lemmas and a particular dimension type only counted once

(3) a. Jacquelyn Sertic (19-23), who threw 213 pitches including 165 strikes in completing both games, retired Oklahoma in order in the bottom of the ninth. That included an over-the-shoulder catch from DeCamp, who *outran the ball* into left center to make the catch. “It says a lot about Jax for her to be able to come back,” Mueller said.

b. He immediately ran away before I could get there myself and *outran me* (I didn’t pick up the elven swiftness skill, usually if I want speed I just mount a beast so running isn’t my characters forte).
Study 1

Do VerbNet-classes allow for higher-level generalizations wrt dimension-based comparisons?

- combinations of *out*-lemmas with concrete information on scalar dimensions
- combinations of verb lemmas and a particular dimension type only counted once

(3) a. Jacquelyn Sertic (19-23), who threw 213 pitches including 165 strikes in completing both games, retired Oklahoma in order in the bottom of the ninth. That included an over-the-shoulder catch from DeCamp, who *outran the ball* into left center to make the catch. “It says a lot about Jax for her to be able to come back,” Mueller said.

b. He immediately ran away before I could get there myself and *outran me* (I didn’t pick up the elven *swiftness skill*, usually if I want *speed* I just mount a beast so running isn’t my characters forte).
Dominant dimensions for verb classes
Cross-listing score

- verb lemmas in VerbNet can belong to more than one class (both homonymy and polysemy)

a) Susan sang to the children.  [MANNER-SPEAKING; syntax: AGENT V \{+DEST.DIR\} RECIPIENT]
b) Sandy sang a song.  [PERFORMANCE; syntax: AGENT V THEME ]
c) The street sang with horns.  [SOUND-EMISSION; syntax: LOCATION V \{with\} THEME ]
Cross-listing score

» *out*-prefixation introduces its own syntactic frame → link to source class unclear

» calculation of a cross-listing score for all lemmas featuring as bases in the out-lemma-dimension combinations

» mean value of additional VerbNet classes per morphological base:
  e.g. *sing*:
    » PERFORMANCE and two additional classes
    » cross-listing score of 2
Cross-listing score

Dimension distribution in % over 7 VerbNet-classes

Cross-listing score:
- PERFORMANCE: 3.1
- RUN: 1.0
- EXIST: 1.1
- SPRAY: 0.7
- SOUND_EMISSION: 2.2
- CARRY: 3.0
- HIT: 3.3
Correlation Dominant dimension-cross-listing score

Pearson’s $r$: -0.83 ($p = 0.02$)
Tokens of *out*-prefixed lemmas and their scalar dimensions

- Again culled from iWeb
- 100 tokens each (if possible) of 4 verb lemmas each from 3 of the VerbNet classes

(4)  
   a. PERFORMANCE: *outdance, outrap, outwrite, outsing*  
   b. RUN: *outrun, outfly, outswim, outsprint*  
   c. EXIST: *outlive, outstay, outwait, outsurvive*
Do clusters of *out*-tokens and dimensions reflect the findings for their particular classes?

- Same general procedure as in study 1
  - Concrete info on dimension in context
  - Here: combinations of *out*-tokens of given lemmas with scalar dimensions
Study 2: PERFORM

Dimension info for 100/all attestations each of 4 PERFORMANCE-verbs

- OUTDANCE
- OUTRAP
- OUTWRITE
- OUTSING

Legend:
- Range
- Intensity
- Loudness
- Quantity
- Duration
- Speed
- Quality
- Unspecified
Study 2: EXIST

Dimension info for 100 attestations each of 4 EXIST-verbs

- OUTLIVE
- OUTWAIT
- OUTSTAY
- OUTSURVIVE

- Duration
- Unspecified
Study 2: RUN

Dimension info for 100 attestations each of 4 RUN-verbs

- **OUTRUN**
- **OUTFLY**
- **OUTSWIM**
- **OUTSPRINT**

Legend:
- **Altitude**
- **Frequency**
- **Maneuverability**
- **Distance**
- **Quality**
- **Speed**
- **Unspecified**
Intermediate conclusion

- VerbNet classes allow for higher-level generalizations wrt dimensions
- divergence from unambiguous dimension distributions possibly explained by numbers of cross-listings, i.e. polysemy/homonymy
- distributions for lemmas largely mirror those for classes
- divergence for some items explained by cross-listing
Part 2

out-derivation in distributional semantics (e.g. Sahlgren 2006):

1. Is the syntactic and dimension-based similarity of out-derivatives reflected in distributional measures?

2. Are the differences in crosslisting scores reflected distributionally?
out-derivation in distributional semantics (e.g. Sahlgren 2006):

- meaning of a linguistic item $\approx$ co-occurrence with other linguistic items
- measures extracted from ukWac (Ferraresi et al. 2008)
- tagged, web-based corpus; 1.9 billion words; British English
- distributional information captured by multidimensional vectors in distributional space
- similarity between items based on cosine similarity
Part 2

1. Is the syntactic and dimension-based similarity of *out*-derivatives reflected in distributional measures?
1. Is the syntactic and dimension-based similarity of *out*-derivatives reflected in distributional measures? Are *out*-derivatives within a class more similar to each other than their bases?
1. Is the syntactic and dimension-based similarity of *out*-derivatives reflected in distributional measures? Are *out*-derivatives within a class more similar to each other than their bases?

2. Are the differences in crosslisting scores reflected distributionally?
1. Is the syntactic and dimension-based similarity of *out*-derivatives reflected in distributional measures? Are *out*-derivatives within a class more similar to each other than their bases?

2. Are the differences in crosslisting scores reflected distributionally? Do *run* and *exist* show similar distributional patterns, and *run* and *perform* less similar patterns?
### Base-to-base similarities

<table>
<thead>
<tr>
<th>class</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>perform</td>
<td>0.18</td>
<td>0.04</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>exist</td>
<td>0.15</td>
<td>0.02</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>run</td>
<td>0.174</td>
<td>0.007</td>
<td>0.164</td>
<td>0.183</td>
</tr>
</tbody>
</table>
## Base-to-base similarities

<table>
<thead>
<tr>
<th>class</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>perform</td>
<td>0.18</td>
<td>0.04</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>exist</td>
<td>0.15</td>
<td>0.02</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>run</td>
<td>0.174</td>
<td>0.007</td>
<td>0.164</td>
<td>0.183</td>
</tr>
</tbody>
</table>
## Base-to-base similarities

<table>
<thead>
<tr>
<th>class</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>perform</td>
<td>0.18</td>
<td>0.04</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>exist</td>
<td>0.15</td>
<td>0.02</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>run</td>
<td>0.174</td>
<td>0.007</td>
<td>0.164</td>
<td>0.183</td>
</tr>
</tbody>
</table>
## Derivation-to-derivation similarities

<table>
<thead>
<tr>
<th>class</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>perform</td>
<td>0.09</td>
<td>0.03</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>exist</td>
<td>0.07</td>
<td>0.02</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>run</td>
<td>0.10</td>
<td>0.03</td>
<td>0.06</td>
<td>0.16</td>
</tr>
</tbody>
</table>
## Derivation-to-derivation similarities

<table>
<thead>
<tr>
<th>class</th>
<th>mean</th>
<th>sd</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>perform</td>
<td>0.09</td>
<td>0.03</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>exist</td>
<td>0.07</td>
<td>0.02</td>
<td>0.05</td>
<td>0.11</td>
</tr>
<tr>
<td>run</td>
<td>0.10</td>
<td>0.03</td>
<td>0.06</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Part 2: further issues

- lexicalization
- specificity: bases themselves already differ in specificity
- general mapping of verbnet-classes against distributional classes
Summary and conclusions

► Part 1
  ▶ Study 1: dominant dimensions for each class, correlation crosslisting-dimension dominance
  ▶ Study 2: class dominant dimension is dominant dimension at lemma level

► Part 2:
  ▶ unexpected results 1: derivation-derivation similarities lowest
  ▶ unexpected results 2: reflection of cross-listing score
Thank you!