

Understanding out-prefixation: merging qualitative and distributional analyses

Martin Schäfer and Sven Kotowski

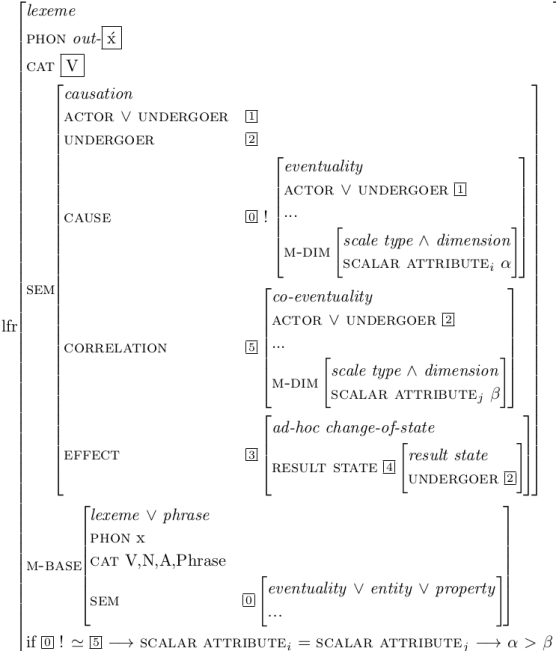
Universität Tübingen and Heinrich-Heine-Universität Düsseldorf

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)

Introduction



$$\boxed{0} ! \left[\begin{array}{l}
 \textit{eventuality} \\
 \text{ACTOR } \vee \text{ UNDERGOER } \boxed{1} \\
 \dots \\
 \text{M-DIM} \left[\begin{array}{l}
 \textit{scale type} \wedge \textit{dimension} \\
 \text{SCALAR ATTRIBUTE}_i \alpha
 \end{array} \right]
 \end{array} \right]$$

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

Study 1

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

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

Study 1

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

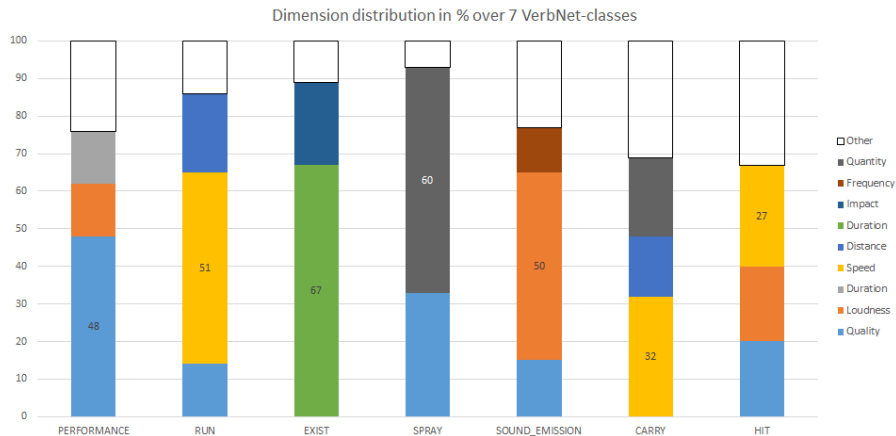
- ▶ combinations of *out*-lemmas with concrete information on scalar dimensions
 - ▶ combinations of verb lemmas and a particular dimension type only counted once
- (3) a. Jacquelyn Sertic (19-23), who threw 213 pitches including 165 strikes in completing both games, retired Oklahoma in order in the 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

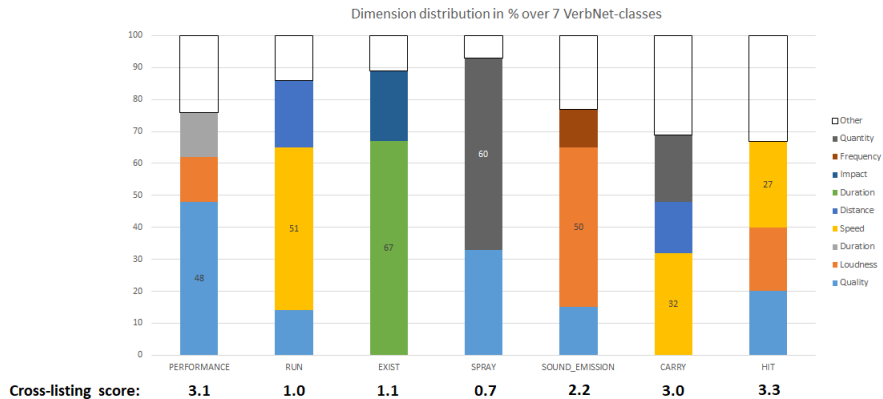
- ▶ verb lemmas in VerbNet can belong to more than one class (both homonymy and polysemy)

a) <i>Susan sang to the children.</i>	[MANNER-SPEAKING;	syntax: AGENT V {+ <i>DEST_DIR</i> } RECIPIENT]
b) <i>Sandy sang a song.</i>	[PERFORMANCE ;	syntax: AGENT V THEME]
c) <i>The street sang with horns.</i>	[SOUND-EMISSION;	syntax: LOCATION V {with} THEME]

Cross-listing score

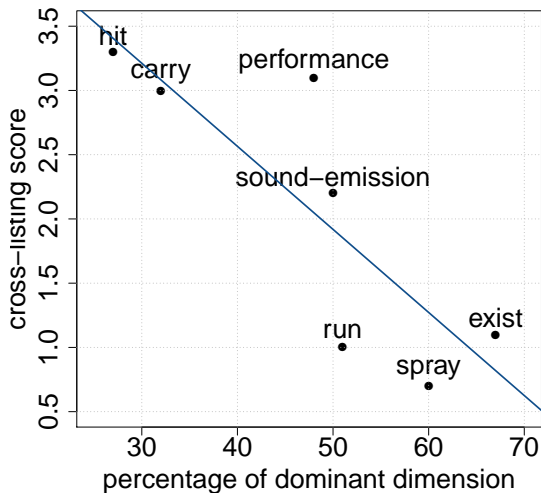
- ▶ *out*-prefixation introduces its own syntactic frame → link to source class unclear
- ▶ calculation of a cross-listing score for all lemmas featuring as bases in the out-lemma-dimension combinations
- ▶ mean value of additional VerbNet classes per morphological base:
e.g. *sing*:
 - ▶ PERFORMANCE and two additional classes
 - ▶ cross-listing score of 2

Cross-listing score



Correlation Dominant dimension-cross-listing score

Pearson's r : -0.83 ($p = 0.02$)



Study 2

Tokens of *out*-prefixed lemmas and their scalar dimensions

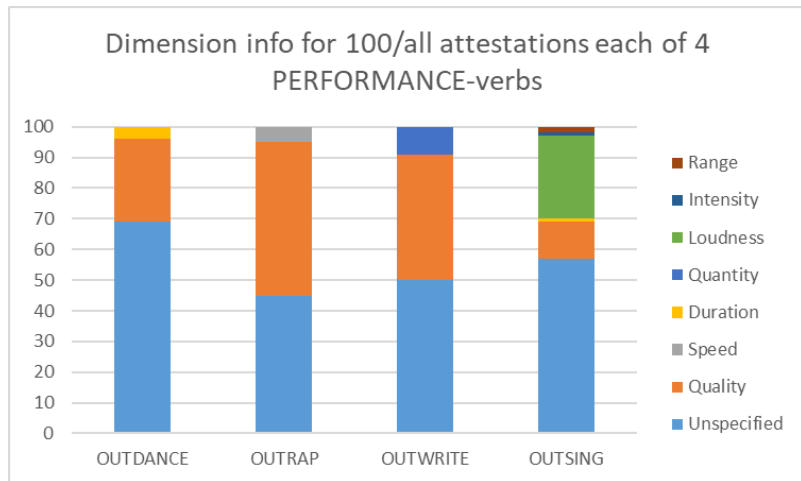
- ▶ Again culled from iWeb
 - ▶ 100 tokens each (if possible) of 4 verb lemmas each from 3 of the VerbNet classes
- (4)
- a. PERFORMANCE: *outdance, outtrap, outwrite, outsing*
 - b. RUN: *outrun, outfly, outswim, outsprint*
 - c. EXIST: *outlive, outstay, outwait, outsurvive*

Study 2 ctd.

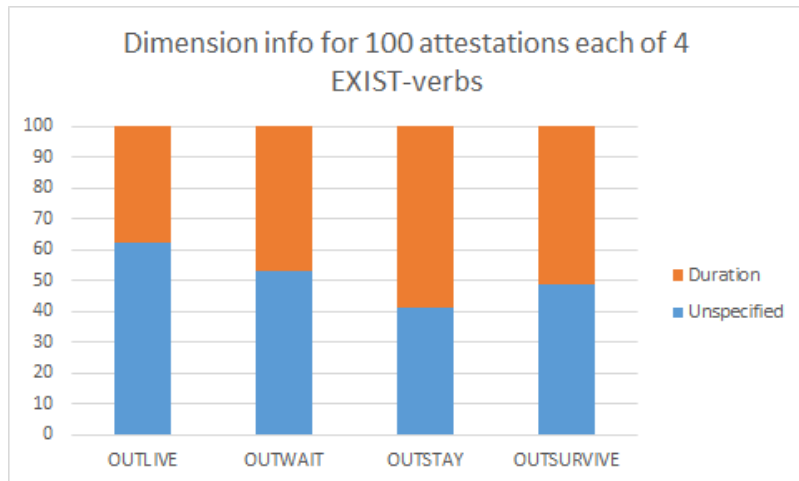
Do clusters of *out*-tokens and dimensions reflect the findings for their particular classes?

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

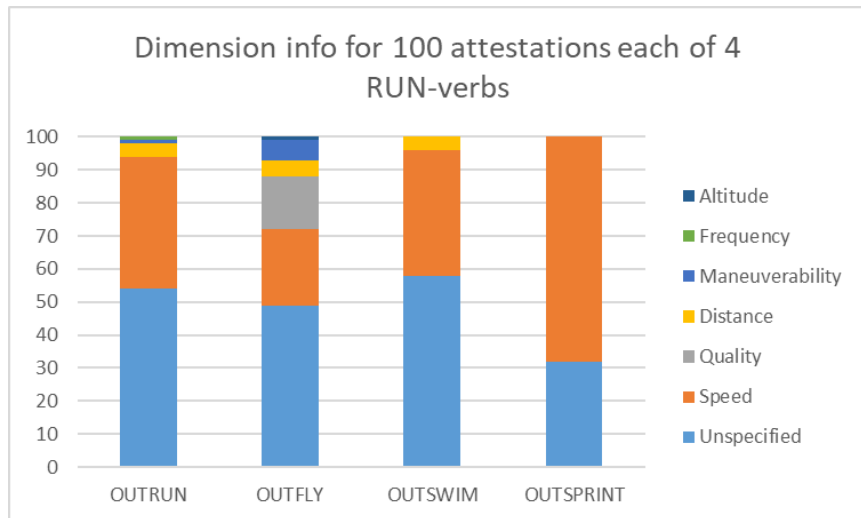
Study 2: PERFORM



Study 2: EXIST



Study 2: RUN



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?

Part 2

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?

Part 2

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?

Part 2

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?

Part 2

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

class	mean	sd	min	max
perform	0.18	0.04	0.13	0.23
exist	0.15	0.02	0.13	0.19
run	0.174	0.007	0.164	0.183

Base-to-base similarities

class	mean	sd	min	max
perform	0.18	0.04	0.13	0.23
exist	0.15	0.02	0.13	0.19
run	0.174	0.007	0.164	0.183

Base-to-base similarities

class	mean	sd	min	max
perform	0.18	0.04	0.13	0.23
exist	0.15	0.02	0.13	0.19
run	0.174	0.007	0.164	0.183

Derivation-to-derivation similarities

class	mean	sd	min	max
perform	0.09	0.03	0.05	0.15
exist	0.07	0.02	0.05	0.11
run	0.10	0.03	0.06	0.16

Derivation-to-derivation similarities

class	mean	sd	min	max
perform	0.09	0.03	0.05	0.15
exist	0.07	0.02	0.05	0.11
run	0.10	0.03	0.06	0.16

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!