Semantic validation of a German derivational lexicon

Britta D. Zeller[*], Sebastian Padó[†], Jan Šnajder[§]

[*]Institute of Computational Linguistics, Heidelberg University
[†]Institute for Natural Language Processing, Stuttgart University
[§]Faculty of Electrical Engineering and Computing, University of Zagreb

25th Meeting of the
International Conference on Computational Linguistics
August 28th, 2014

Derivational lexicons

- Derivation: Morphological word formation process; basis and derived word share the stem: *to sleep* $\rightarrow$ *sleepy*
- Cluster derivationally related lemmas into derivational families: *to sleep*$_V$ – *sleepy*$_A$ – *sleepless*$_A$ – *sleep*$_N$ – *sleeping*$_A$ – ...
  - Set of morphologically related lemmas across POS
  - Derivational lexicons: CatVar [Habash and Dorr, 2003]; DErivBase [Zeller et al., 2013]
- Assumption for use in NLP: Derivational families capture semantic relatedness across POS boundaries:
  - Textual Entailment [Szpektor and Dagan, 2008]: *the losing*$_A$ $X \leftrightarrow X$ *loses*$_V$
  - Smoothing distributional models [Padó et al., 2013]: $sim(oldish, ancient) = 0$ $sim(old, ancient) > 0$
    $$sim((oldish \circ old), ancient) > 0$$

Derivational lexicons

- Derivation: Morphological word formation process; basis and derived word share the stem: *to sleep* $\rightarrow$ *sleepy*
- Cluster derivationally related lemmas into derivational families: $\textit{to sleep}_V$ – $\textit{sleepy}_A$ – $\textit{sleepless}_A$ – $\textit{sleep}_N$ – $\textit{sleeping}_A$ – ...
  - Set of morphologically related lemmas across POS
  - Derivational lexicons: CatVar [Habash and Dorr, 2003]; DErivBase [Zeller et al., 2013]
- Assumption for use in NLP: Derivational families capture semantic relatedness across POS boundaries:
  - Textual Entailment [Szpektor and Dagan, 2008]: $\textit{the losing}_A\ X \leftrightarrow X\ \textit{loses}_V$
  - Smoothing distributional models [Padó et al., 2013]: $sim(oldish, ancient) = 0$ $\qquad sim(old, ancient) > 0$

$$sim((oldish \circ old), ancient) > 0$$

But: Does membership in the same derivational family already indicate high semantic relatedness?

Derivational lexicons: Limits

- Morphological relatedness implies semantic relatedness often, but not always
- Semantic dissimilarity possible through, e.g.:
    - Diachronic changes in word meaning: *Knie*$_N$ - *beknien*$_V$ *(knee - to beg)*
    - Meaning-changing derivations: *eitel*$_A$ - *vereiteln*$_V$ *(vain - to block)*

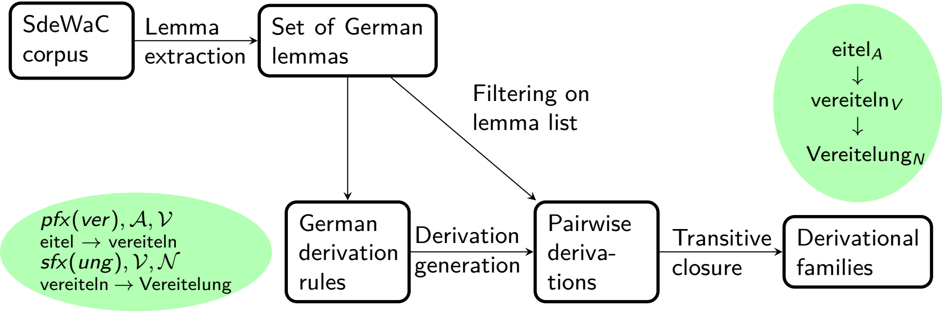
Derivational lexicons: Limits

- Morphological relatedness implies semantic relatedness often, but not always
- Semantic dissimilarity possible through, e.g.:
  - Diachronic changes in word meaning: *Knie*$_N$ *- beknien*$_V$ *(knee - to beg)*
  - Meaning-changing derivations: *eitel*$_A$ *- vereiteln*$_V$ *(vain - to block)*

⇒ Goal: Semantic validation of DErivBase

Methods to determine, for each derivationally related lemma pair, whether it is in fact semantically related

Rule-based induction of DErivBase [Zeller et al., 2013]



- 267 derivation rules
- ∼69K lemmas grouped into ∼17K derivational families

Evaluation of DErivBase [Zeller et al., 2013]

- 4,000 lemma pairs with five annotation classes:
  - **S** – morphologically and **S**emantically related
    Speicher$_N$ – *speichern*$_V$ (*storage*$_N$ – *to store*$_V$)
  - **M** – only **M**orphologically related
    bomben$_V$ – *bombig*$_A$ (*to bomb*$_V$ – *smashing*$_A$)
  - **N** – **N**o morphological relation
    Säge$_N$ – *Sage*$_N$ (*saw*$_N$ – *legend*$_N$)
  - **L** – **L**emmatization error
    Haufe$_N$ – *Häufung*$_N$ (*N/A* – *accumulation*$_N$)
  - **C** – **C**ompound relation
    filmen$_V$ – *Filmende*$_N$ (*to film*$_V$ – *end of film*$_N$)
- Inter annotator agreement: $\kappa > 0.7$
- Measures: Recall, Precision, with respect to positive class
  - In [Zeller et al., 2013], both **S and M** counted as positive class
  - Accepting only **S** as correct: Recall: 93.8%; Precision: 76.7%

$\Rightarrow$ Semantic validation important to improve lexicon precision

Semantic validation

- Binary classification task: Decide for a lemma pair of the same derivational family: Semantically related or not (**S** vs. non-**S**)?
  - No prediction of whole families, but pairs drawn from them
  - Pair information: lemma_1, lemma_2, connecting rule path
  - "Simplex" and "complex" paths: $\textit{eitel}_A \rightarrow \textit{vereiteln}_V \rightarrow \textit{Vereitelung}_N$
- Information sources: Distributional semantics, derivation rules
- Two basic hypotheses:
  - **Hypothesis 1**: High distributional similarity between derivationally related words indicates semantic relatedness.
  - **Hypothesis 2**: Derivation rules differ in their reliability.
- Data analyses on each information source, implementation of findings into ML classification model

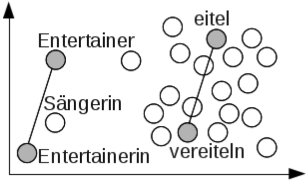
Dataset

- Re-use pairs of annotation in [Zeller et al., 2013] relevant for v1.4.1
- Total: 2,543 lemma pairs
- $\sim$ 75% are **S** pairs; very good majority class baseline
- 70:30 split into development (1,780 pairs) and test set (763 pairs)
- Development set: Basis for our analyses

- Distributional similarity model: Large German web corpus [Faaß et al., 2010], standard BOW model

Analysis I: Distributional rank outperforms raw cosine

- Finding: Raw cosine is inadequate:
  - Infrequent words of DErivBase unreliably represented
  - Conceptual issues, like markedness: $cos(Entertainer, Entertainerin) = 0.1$
- Instead: Semantic similarity in terms of ranks [Hare et al., 2009], [Lapesa and Evert, 2013]: Take density into account



$dist(Entertainer, Entertainerin) = dist(eitel, vereiteln)$

$rank(Entertainer, Entertainerin) = 3$
$rank(eitel, vereiteln) = 16$

⇒ **Hypothesis 1'**: High *rank-based* distributional similarity between derivationally related words indicates semantic relatedness.

Analysis II: Derivation rules

1. **Hypothesis 2** confirmed: Not all derivation rules are meaning-preserving
   - Semantic dissimilarity between the two lemmas; e.g., prefixation derivations:
     *hören$_V$ *- aufhören$_V$ (to listen *- to stop)*: **M**
   - Mostly meaning-preserving; e.g., definition of specific semantic aspects:
     Entertainer$_N$ - Entertainerin$_N$: **S**

2. Lemmas linked by a complex path show a "weakest link" behaviour
   - One meaning-changing rule in a path is enough to cause overall dissimilarity

eitel$_A$
$\downarrow$ **M**
vereiteln$_V$
$\downarrow$ **S**
Vereitelung$_N$

Features

Implement features for observed aspects of both analyses, combine them into a Machine Learning classifier

Three feature groups (34 features total):

- Distributional features (6):
  Absolute cosine similarity, cosine rank similarity, ...
- Rule-based features (25):
  Rule reliability measures, path length, ...
  Due to "weakest link" behaviour: Choose most pessimistic value for pairs with complex paths
- Hybrid features (3):
  Combination of distributional and derivation rule information

Classification model

- Binary decision: **S** vs. non-**S** (**M**, **N**, **L**, **C**)
- Nonlinear model: Radial Basis Function (RBF) kernel in LIBSVM
- Training with 3-fold cross-validation

Results of various feature combinations

Validation method	Precision	Recall	F_1	Accuracy
Majority baseline (**S**)	72.6	**100**	84.1	72.6

Results of various feature combinations

Validation method	Precision	Recall	F_1	Accuracy
Majority baseline (**S**)	72.6	**100**	84.1	72.6
Classifier, *only "cosine similarity" feature*	72.6	100	84.1	72.6
Classifier, *only "similarity rank" feature*	**80.3**	90.3	**85.0**	**76.8**

- **Hypothesis 1'** ✓: Rank-based sim. more suitable than raw cosine

Results of various feature combinations

Validation method	Precision	Recall	F_1	Accuracy
Majority baseline (**S**)	72.6	**100**	84.1	72.6
Classifier, *only "cosine similarity" feature*	72.6	100	84.1	72.6
Classifier, *only "similarity rank" feature*	80.3	90.3	85.0	76.8
Classifier, *distributional group*	80.5	96.6	**87.8**	80.5
Classifier, *rule-based group*	**82.7**	93.1	87.6	**80.9**
Classifier, *hybrid group*	80.4	95.3	87.2	79.7

- **Hypothesis 1'** ✓: Rank-based sim. more suitable than raw cosine
- **Hypothesis 2** ✓: Rule-based features contribute esp. in precision

Results of various feature combinations

Validation method	Precision	Recall	F_1	Accuracy
Majority baseline (**S**)	72.6	**100**	84.1	72.6
Classifier, *only "cosine similarity" feature*	72.6	100	84.1	72.6
Classifier, *only "similarity rank" feature*	80.3	90.3	85.0	76.8
Classifier, *distributional group*	80.5	96.6	87.8	80.5
Classifier, *rule-based group*	82.7	93.1	87.6	80.9
Classifier, *hybrid group*	80.4	95.3	87.2	79.7
Classifier, *all features*	**86.2**	93.9	**89.9**	**84.7**

- **Hypothesis 1'** ✓: Rank-based sim. more suitable than raw cosine
- **Hypothesis 2** ✓: Rule-based features contribute esp. in precision
- Feature groups are complementary: Combination performs best
- Overall improvement: >13% in precision, >5% in F_1

Related work

Little work about semantic validation of derivational lexicons

- [Jacquemin, 2010]: Semantic validation of a French derivational lexicon [Gaussier, 1999], requiring elaborate dictionary information

Learning morphology from distributional models:

- Unsupervised morphology induction [Schone and Jurafsky, 2000, Baroni et al., 2002]
- Induction of semantic classes [Boleda et al., 2012, Schulte im Walde, 2006]
- Opposite direction: Use derivational morphology to improve distributional models [Luong et al., 2013, Lazaridou et al., 2013]

Summary

- Derivational lexicons deliver semantic information, but it is mingled with purely morphological information
- We made a step towards semantic validation of such lexicons by combining distributional and derivation rule-based information
- Major findings:
  - Particularities of derivational relationships influence performance of distributional methods
    - Cosine similarity is not a good indicator, but rank-based similarity is
  - Rule-based features are simple, yet effective for semantic validation
  - Both sources contribute complementary information

Thank you for your attention.

Baroni, M., Matiasek, J., and Trost, H. (2002).
Unsupervised Discovery of Morphologically Related Words Based on Orthographic and Semantic Similarity.
Computing Research Repository, cs.CL/0205006.

Boleda, G., Schulte im Walde, S., and Badia, T. (2012).
Modeling Regular Polysemy: A Study on the Semantic Classification of Catalan Adjectives.
Computational Linguistics, 38(3):575–616.

Faaß, G., Heid, U., and Schmid, H. (2010).
Design and application of a gold standard for morphological analysis: SMOR in validation.
In Proceedings of the Seventh International Conference on Language Resources and Evaluation, pages 803–810.

Gaussier, E. (1999).
Unsupervised learning of derivational morphology from inflectional lexicons.
In ACL Workshop Proceedings on Unsupervised Learning in Natural Language Processing, pages 24–30, College Park, Maryland.

Habash, N. and Dorr, B. (2003).
A categorial variation database for English.
In Proceedings of the Anuual Meeting of the North American Association for Computational Linguistics, pages 96–102, Edmonton, Canada.

Hare, M., Jones, M., Thomson, C., Kelly, S., and McRae, K. (2009).
Activating Event Knowledge.
Cognition, 111(2):151–167.

Jacquemin, B. (2010).
A derivational rephrasing experiment for question answering.
In Proceedings of the Conference on Language Resources and Evaluation, pages 2380–2387, Valletta, Malta.

Lapesa, G. and Evert, S. (2013).
Evaluating neighbor rank and distance measures as predictors of semantic priming.
In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 66–74, Sofia, Bulgaria.

Lazaridou, A., Marelli, M., Zamparelli, R., and Baroni, M. (2013).
Compositional-ly derived representations of morphologically complex words in distributional semantics.
In Proceedings of the Association for Computational Linguistics, pages 1517–1526, Sofia, Bulgaria.

Luong, M.-T., Socher, R., and Manning, C. D. (2013).
Better word representations with recursive neural networks for morphology.
In Proceedings of the Conference on Natural Language Learning, pages 104–113, Sofia, Bulgaria.

Padó, S., Šnajder, J., and Zeller, B. (2013).
Derivational smoothing for syntactic distributional semantics.
In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 731–735, Sofia, Bulgaria.

Schone, P. and Jurafsky, D. (2000).
Knowledge-free induction of morphology using latent semantic analysis.
In Proceedings of the Conference on Natural Language Learning, pages 67–72. Lisbon, Portugal.

Schulte im Walde, S. (2006).
Experiments on the Automatic Induction of German Semantic Verb Classes.
Computational Linguistics, 32(2):159–194.

Szpektor, I. and Dagan, I. (2008).
Learning Entailment Rules for Unary Templates.
In Proceedings of the 22nd International Conference on Computational Linguistics, pages 849–856, Manchester, UK.

Zeller, B., Šnajder, J., and Padó, S. (2013).
DErivBase: Inducing and evaluating a derivational morphology resource for German.
In Proceedings of ACL, Sofia, Bulgaria.