The Collection
Semantically related English word pairs, rated for the strength of the semantic relation holding between them
Part of a larger project, whose goal is to characterize paradigmatic relations cross-linguistically
- German: (Scheible and Schulte im Walde, 2014). IMS Stuttgart
- Italian: Computational Linguistics Lab, University of Pisa (collection ongoing)

paradigmatic relations: synonymy, antonymy, hypernymy

3 parts of speech: nouns, verbs, adjectives

degrees of relatedness: Target: artist
Relation: synonymy
painter (strongly related)
creator (weakly related)
scientist (negatively related).

Step 1: Generation Experiment
On Amazon Mechanical Turk, native speakers were asked to generate related words for 99 English targets per part-of-speech
Random selection of targets from WordNet (Miller, 1995) with a stratified sampling technique (Scheible and Schulte im Walde, 2014). Criteria:
- polysemy class: I) one sense; II) two senses; III) > 3 senses
- frequency classes: I) low (200–2,999); II) mid (3,000–9,999); III) high (>10,000)
- size of the WordNet semantic class

Step 2: Rating Experiment
Objective: find pairs selected from the generated data
Goal: find pairs for which a full tableau was available. Criteria:
- at least 2 different relateda had been generated
- a strongly related word (e.g., painter) was produced at least 4 times
- a weakly related word (e.g., creator) was produced twice or once
- a negatively related word was produced at least twice for the opposite relation: ANT for SYN and HYP, SYN for ANT (e.g., painter, antonym, scientist)

285 target, relation, relatum, related word / direction combinations

Ratings collected with AMT:
✓ Do you think that the following two words are synonyms?
✓ 6 points scale (0–5)
✓ 10 workers per target, relation, relatum / triple, per each direction

The Resource

Case Study: Directionality
Are some relations/parts-of-speech more asymmetric than others?
Method: item-based prediction with linear regression models

References & Acknowledgments

We thank Giulia Benotto and Alessandro Lenci for providing us the generation data. This research was funded by the DFG Collaborative Research Centre SFB 732 (Gabriella Lapesa) and the DFG Heisenberg Fellowship SCHU-2580/1-1 (Sabine Schulte im Walde).