## Introduction

### Goal
Providing lexicographers / language learners with salient semantic relation instances for a given concept.

(e.g. dog: tail vs. heart)

### Current Focus
Composite part relations: [adj,modifier] [noun]

## Supporting the Idea

### Analysis results of our preceding feature production experiment for DE + IT speakers (collection of descriptions for concrete concepts):
Modifiers used further specified a feature – in contrast to twoword expressions.

(e.g. dog: wet nose)

### Further Motivation
Previous automatic extraction algorithms have been concentrating on atomic relation instances – but lexicographic projects make use of composites, e.g. WordNet (~ 1300 instances), ConceptNet (~ 68000 instances).

## Method

### Starting from DE data from production experiment – concepts with produced parts– look up corpus frequencies for pairs [adj]...[noun], separated maximally by 3 words:

- Within wide context of concept (20 sentences) – concept-specific, but sparsity problem,
  e.g. bear: white fur, but not: broom: long bristles
- Not considering concept context – more exhaustive, but modifiers not concept-specific,
  e.g. pineapple: thick peel, but owl: yellow eye on low rank

\[ \text{Combine information: } \text{freq}_{\text{comb}} = \text{freq}_{\text{mod}} \cdot \text{part}_\text{in concept context} \times \text{freq}_{\text{mod}} \cdot \text{part}_\text{not considering concept} \]

Select top 5 candidates from each of these ranked lists.

## Performance

### Gold Standard: Composite parts of production experiment (DE)

Method trained on DE data performs similarly well on IT data.

\[ \text{Cosine distance reranking: } \text{freq}_{\text{comb}} \times (1 + \text{cos-dist}) \]

Raise combined freq value according to similarities to top 100 contextless modifiers – distance calculation based on number of coocurrences with nouns.

## Performance on New Concepts

Select DE modifiers for the concept-part pairs of McRae et al.’s production experiment.

### Gold Standard (from judgement experiment):
Modifiers with acceptance rate \( \geq 0.75 \)

## Production vs. Perception

### Acceptance of those DE concept–part–modifier triples that were collected in the production experiment:
Small overlap of modifiers that were produced and accepted (46), compared to the numbers of modifiers exclusively produced (53) or accepted (42),

Distribution of acceptance rates only for the triples from the production experiment compared to the overall distribution

T-test: \( p=0.51 \)

## Conclusion

- What is produced quite differs from what is perceived as salient
- Our method yields modifiers of both kinds

### Visual Exploration

Only produced:

- monkey: short leg, and also: bear: white fur

Only accepted:

- corn: large leaf, but also: bear: white tooth

## Next in Line

- Resulting salient modifiers useful for identifying salient parts?
- Adaptability of method to other semantic relation types, and with possible further improvement?

## References