The Alan Turing Institute

Polysemy Patterns

in Human Judgements and Contextualised Language Models

Janosch Haber

DALI Project
Cognitive Science Group (CogSci)
Queen Mary University of London

Queen Mary University of London

DALI Project Overview (2018-2021)

- **2018: Reference and Common Ground**
  Finalising the PhotoBook project, a large-scale dataset for multi-turn, visually-grounded dialogue collected during my Master's at the University of Amsterdam. *Presented at ACL 2019*

- **2019: Under-specification in Dialogue**
  Pilot on classifying under-specified and ambiguous expressions in dialogue. *Presented at SemDial 2020*

- **2020 and 2021: Word Sense Similarity**
  Collecting and analysing a human-annotated dataset rating the similarity of different interpretations of lexically ambiguous expressions, and investigating contextualised language models to predict these ratings. *Presented at PaM 2020, *Sem 2020 and Findings of EMNLP 2021*

Ignorance is Bliss

- **TL;DR**
  In my research project I think about language use meant to NOT be thought about. Instead of asking people for their intuitions, large-scale language resources and language models could be used to learn more about how we process language.

- **Under-specification**
  Language develops through evolution, based on its use. One of the main factors of language efficiency is under-specification: We don't need to clarify everything to 100% to still reach our communicative goals.

- **Warning**
  Thinking about under-specification can mess with your (perfectly fine) intuition about what is correct language use.

Under-specification and Lexical Ambiguity

- **The mental lexicon**
  In order to process language, we need some form of mental lexicon to specify the meaning of words
  - Some words are lexically ambiguous, which means that they can have different interpretations in different contexts

- **How are (lexically ambiguous) words represented in the mental lexicon?**

- **Disagreeing data**
  Studies on lexical ambiguity present disagreeing results. I believe that one reason for this is oversimplifying the phenomenon

Lexical Ambiguity

- **Lexical ambiguity as a spectrum**

- **Context, context, context**
  Almost all content words can (slightly) change their meaning based on the context they are used in

- **Inherent meaning**
  Some words come with a set of inherent interpretations that can be elicited by specific contexts

Polysemes

- **Homonyms are words with multiple (unrelated) meanings**
  - The match burned my fingers. [lighter]
  - The match ended without a winner. [sports game]
- **Polysemes are words with multiple distinct but related senses**
  - The school is on fire. [building]
  - The school banned light-up trainers. [leadership team]
  - The school visited the new opera. [group of pupils]
  - The school called Tom's parents. [receptionist]
- **Similarity is only superficial**
  Homonyms need disambiguation, polysemes usually can be left under-specified

Processing Differences

- The bank

Processing Differences

- **The bank sloping down to the river was overgrown with weeds.**

- **Instant Interpretation**
  Homonym interpretations are instant – and specific. If the wrong interpretation was chosen initially, it needs to be corrected. You probably noticed this correction when reading the sample sentence – and it can be measured in experiments.

- Polysemes usually don't inflict any processing costs. You're usually not even aware that you encountered one. And that's the point.

- **Different lexicon entries**
  Common hypothesis: Homonyms have different entries in the lexicon, polysemes are merged into a single under-specified entry

Co-predication

- **Co-Predication tests**
  Traditionally, co-predication tests are used to distinguish homonyms and polysemes
  - # The match burned my fingers and ended without a winner.
  - Lunch was delicious but took forever.

- **Spoiler: It's not that simple**
  Some polysemes lead to unacceptable or zeugmatic co-predication
  - # The newspaper fired its editor in chief and got wet in the rain.
  - # They took the door off its hinges and walked through it.

Cutting through the Fog

- **Hypothesis: Polysemy is not as simple a phenomenon as often assumed**

- **Questions, questions, questions**
  - Does polysemy form a spectrum, too?
  - Are there systematic patterns in polysemy?
  - Can we establish a taxonomy of polyemes? A taxonomy of their senses?
  - Can we measure the “distance” between two interpretations?
  - Does the distance between senses correlate with their co-predication acceptability?

  - Can we infer word sense distance from corpora of language use?
  - Can language models predict word sense distance?
  - Can we use computational tools to aid the linguistics research?

Experiments

- **Goal 1: Collect empirical data about word sense similarity that allows for a systematic evaluation of polysemy patterns**

- **Crowdsourcing on Amazon Mechanical Turk**
  We collected close to 20,000 judgements on word sense similarity in sentence pairs (explicit) or co-predication structures (implicit).

- **Goal 2: Investigate how (contextualised) language models represent word sense**

- **Welcome to Sesame Street!**
  Recent models like ELMo and BERT can produce an encoding of a specific word in a specific context based on immense amounts of training data. The result is a vector that by itself is meaningless to us. But we can calculate the similarity of the vectors in different contexts.

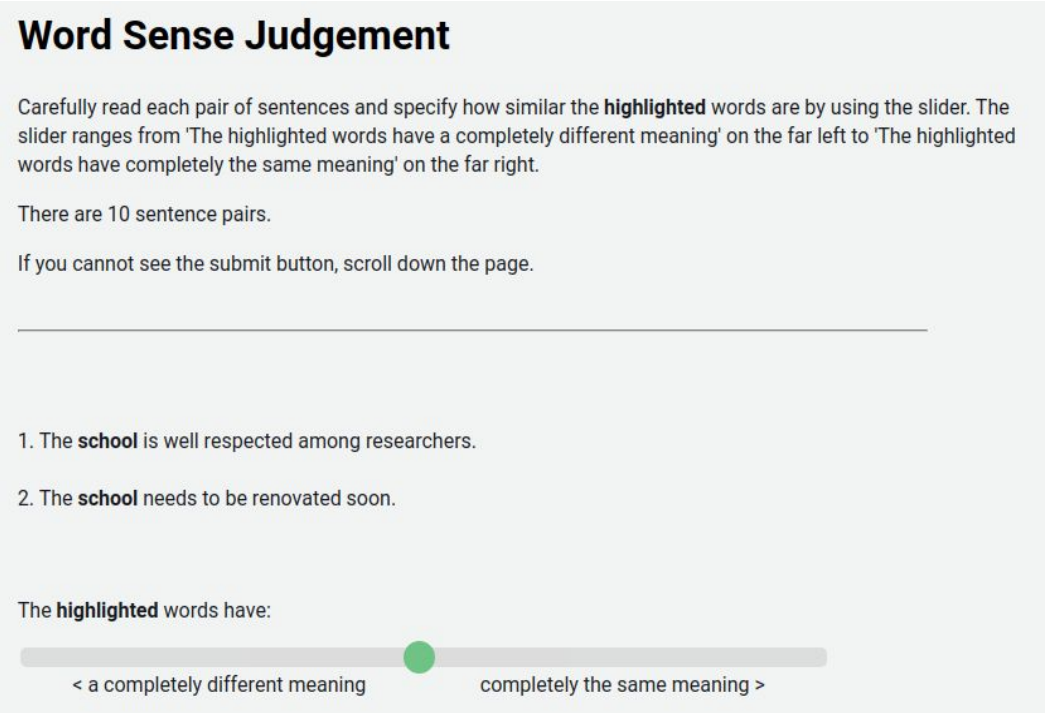
Strict Working Definitions

- **Clear Definitions**
  By clearly defining the (sub-)type of polysemic alternation we investigate, we hope to get meaningful, clear results - from which we might generalise

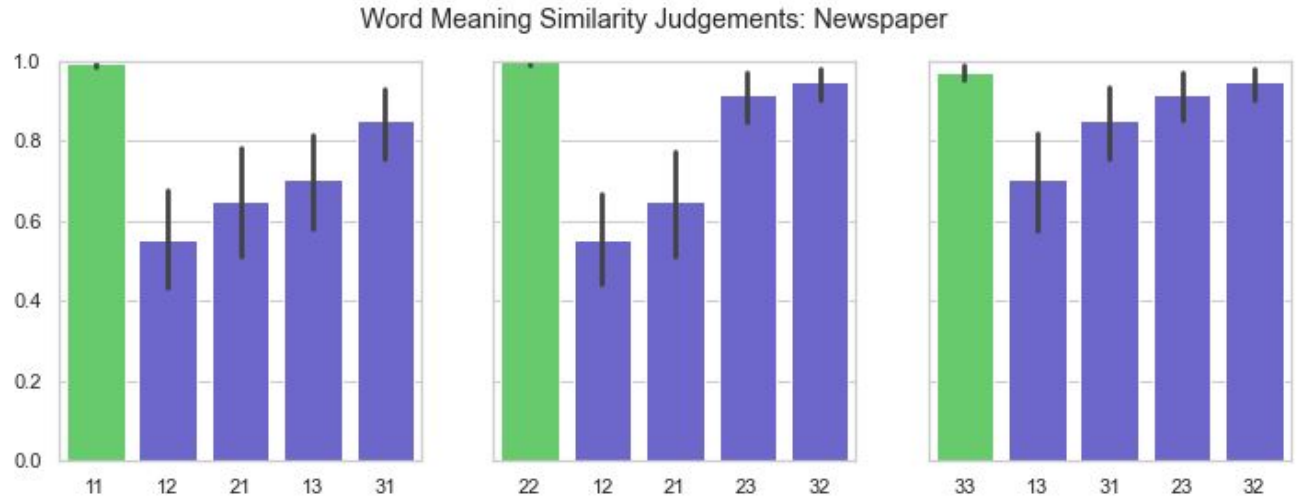
- **Regular, metonymic polysemy**
  For our study, we focus on ten different types of regular (systematic) metonymic polysemic nouns.
  - Metonymy: multiplicity of sense where all senses refer to different aspects/facets of the same concept.
  - Regularity: different word forms allow for the same (sub-)set of alternations.
  - Nouns: You know what nouns are. For now we exclude proper nouns

- Samples are highly controlled and generated through a custom template

Data Collection

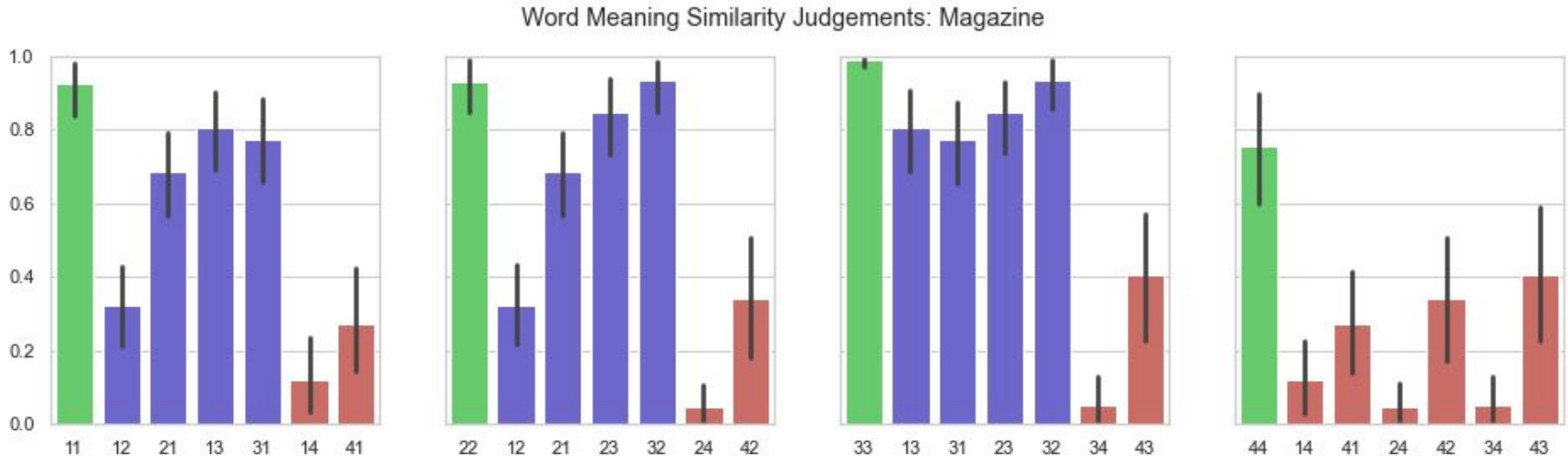


Significant Similarity Differences



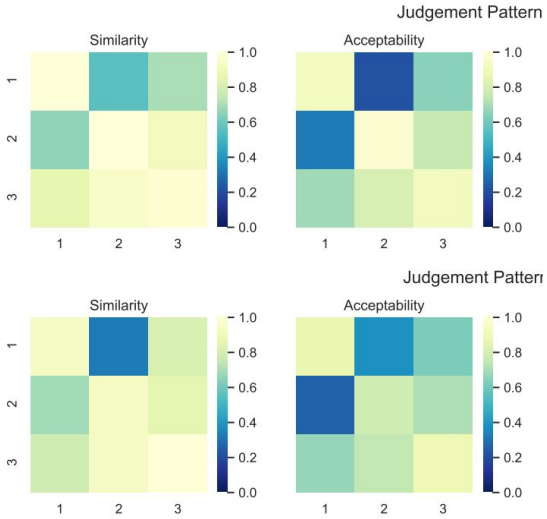
- Some cross-sense polysemic sense combinations receive a significantly lower similarity rating than same-sense pairings - and sometimes also other cross-sense readings

How low can you go?



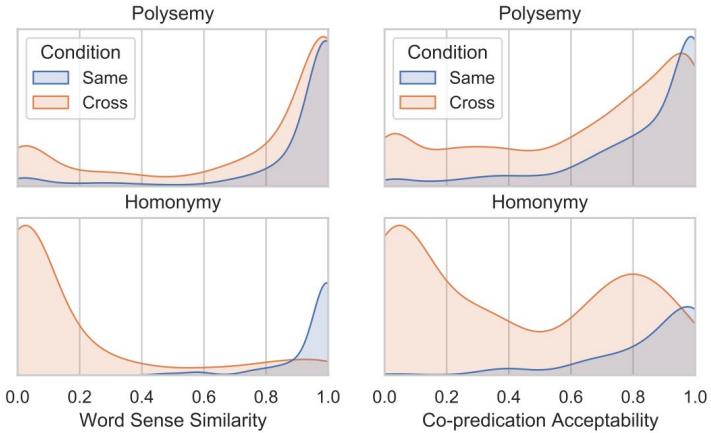
- In some cases, the similarity ratings for cross-sense polyseme pairs are as low - or lower than those for cross-sense homonym samples.

Patterns of Polysemy



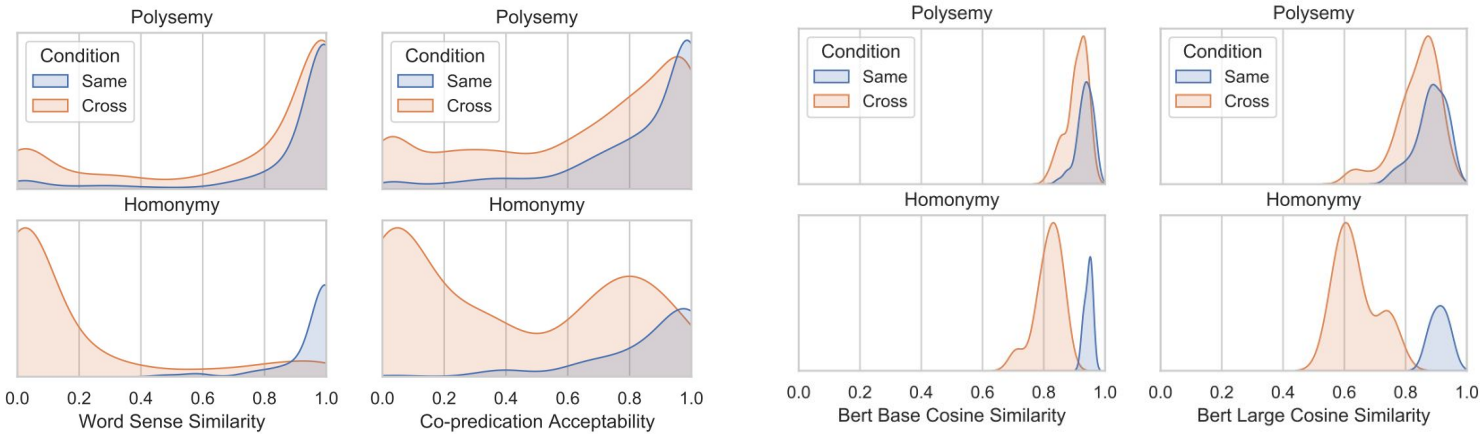
- **Systematic patterns**
  For some types of polysemic alternation, the similarities between senses form consistent patterns across different targets.

Graded Word Sense Similarity



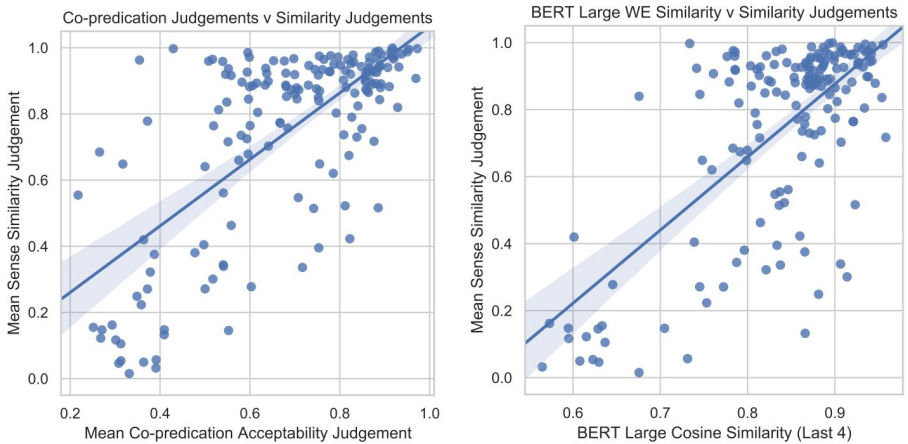
- **Spectrum of Interpretations**
  Empirically, word sense similarity and co-predication acceptability seem to form a spectrum. Co-predication acceptability is more coarse and prone to noise.

Contextualised Language Models



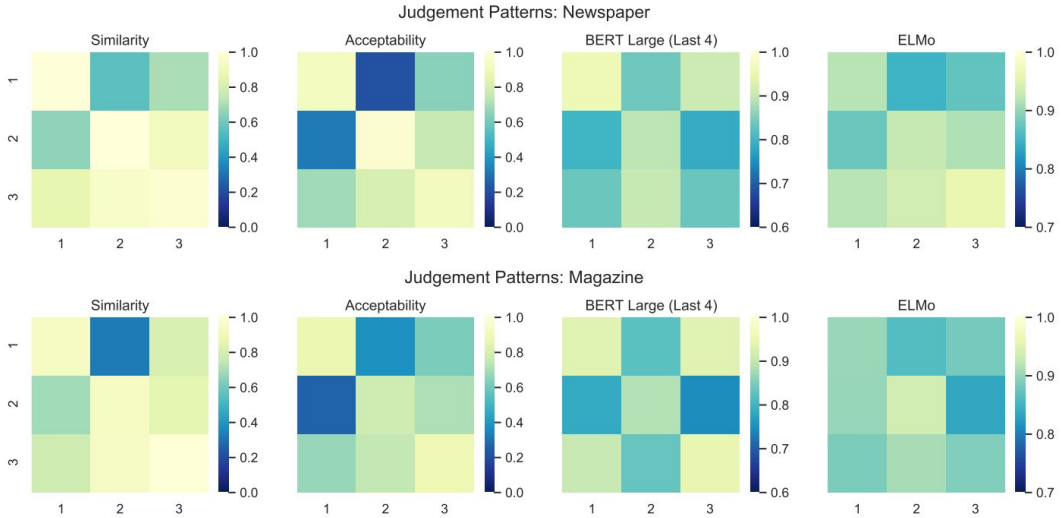
- **Context-sensitivity**
  Contextualised word embeddings can distinguish homonyms from polysemes based on the similarity of use.

BERT Large



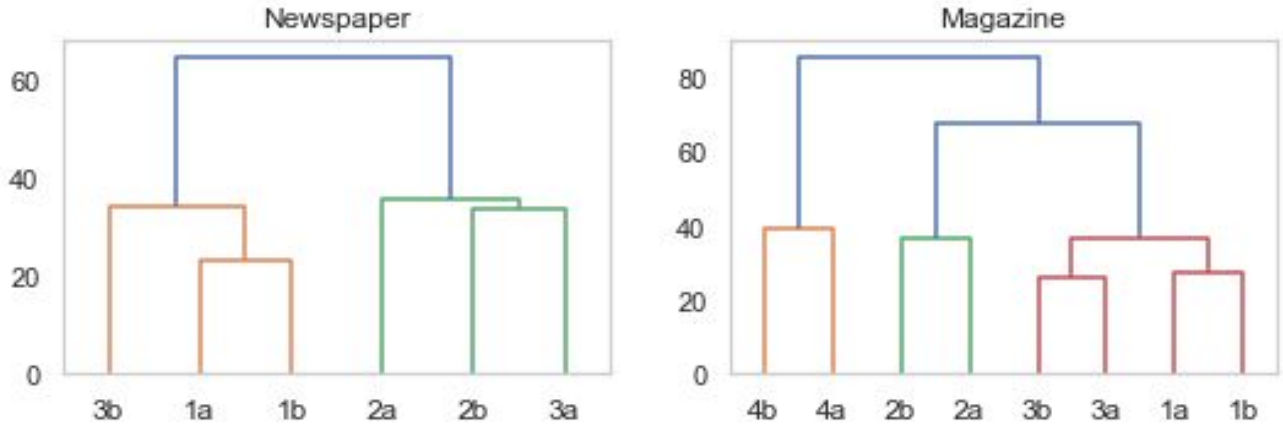
- **Predictive Power**
  BERT Large predicts word sense similarity to the same degree as co-predication acceptability does (Pearson's r = 0.7)

Patterns of Polysemy



- **Systematic patterns**
  For some types of polysemic alternation, the similarities between senses form consistent patterns across different targets.

Clustering of Senses



- **Hierarchical sense clustering**
  Simply based on embeddings cosine similarity, BERT Large can cluster different senses of some types of alternations,

The Alan Turing Institute

Thanks!

Janosch Haber
DALI Project / Queen Mary University of London

j.haber@qmul.ac.uk / janoschhaber.com

Disagreements and Language Interpretation (DALI) / ERC grant number 695662
Led by Prof. Dr. Massimo Poesio

Queen Mary University of London