Reading Disambiguation of newly derived nominalizations in context: A Distributional Semantics approach (2018)

Gabriella Lapesa[1] Lea Kawaletz[2] Ingo Plag[2]
Marios Andreou[2] Max Kisselew[1] Sebastian Padó[1]

[1]University of Stuttgart, Germany

[2]Heinrich-Heine-Universität Düsseldorf, Germany

Alice Missud, *Experimental and computational approaches to morphology and the lexicon*, November 8 2019

Research questions

- How can we assess the semantic outputs of a productive deverbal suffix?
- How do speakers disambiguate the semantic interpretation of newly derived words in context?
- **Are distributional semantics able to tackle these questions?**

Why did I choose this paper?

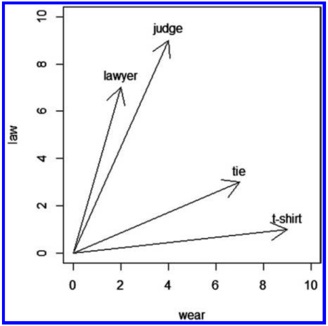
- Few papers have explored word formation using distributional semantics, especially when it comes to neologisms
- Useful for describing the semantic outputs of a specific productive suffix compared to others:
  - French: -age / -ment / -ion / conversion
  - English: -ing / -ment / -ion / conversion
- Interesting to know how much context is enough to disambiguate the semantic interpretation of a newly derived word

Distributional semantics

Foundational assumption: "You shall know a word by the company it keeps" (Firth, 1957:11)

Table 3: Matrix with frequencies of co-occurrence for four target words

Context \ Target	*t-shirt*	*tie*	*lawyer*	*judge*
wear	9	7	2	4
law	1	3	7	9



Previous work

- Kisselew et al. (2016) used distributional semantics to predict the diachronic direction of conversion (verb-noun)
- Cotterell & Schütze (2017) use semantic vectors for morphological segmentation
- Varvara's thesis (2017) explored the semantic differences between rival suffixes using word embeddings
- Not mentioned: Wauquier et al. (2018) tried to measure how close French *-age* / *-ment* / *-ion* derivatives are in terms of distributional properties using barycentric vectors

First paper to use distributional semantics on **newly derived words**.

Hypotheses

- Immediate context helps disambiguating the reading of a neologism
- Semantic classes (event, product, result, etc) have distinct distributional properties
- Distributional properties of low-frequency words can be investigated using quantitative tools

Task

Goal: classify the semantic interpretation of new *-ment* derivatives using word vectors.

- Focus on *-ment* derivatives. *-ment* is an old but still productive suffix that was borrowed from French.
  Multiple readings available:
  - Event: *assessment*
  - Result: *improvement*
  - State: *contentment*
  - Product: *pavement*
  - Instrument: *refreshment*
  - Location: *embankment*
  - Patient/theme: *investment*
- Focus on neologisms

Task

Goal: classify the semantic interpretation of new *-ment* derivatives using word vectors.

- Focus on *-ment* derivatives
- Focus on neologisms
  - Long established derivatives are highly lexicalized and can often be semantically opaque.
  - Newly derived words are rare $\rightarrow$ speakers do not have stored available meanings for these forms.
  - Plag (1999): they may be transparent in order to enable successful communication.
  - Context can be investigated with little bias.

Data

Finding "ambiguous" neologisms

- Oxford English Dictionnary (OED): contains dated citations (18 derivatives found with their context)
- COCA corpus to extract words with frequency of 1 to 3:
  - for each verbal base in VerbNet, searched V + -ment (117 found)
  - any hapax that is not in VerbNet (60 found)
- Only kept nouns derived on psych, change of state, putting and force verbs (Levin's classification, 1993): **55 derivatives**
- WebCorp (Renouf et al. 2006), GloWbE (Davies, 2013) and Google to find attestations: 406 contexts found, covering all possible interpretations.

Categorization

Attestations were annotated manually by three linguists (inter-speaker agreement of at least 2/3).

- EVENTIVE: all subtypes of events, processes or states.
- NON-EVENTIVE: concrete entities (objects, animals, persons) and abstract entities (quantities or means of communication).
- AMBIGUOUS: both interpretations.
- UNCLEAR: ?

They can also be abstract (idea, quality, state, event) or concrete (exist in a physical form), and ambiguous if both interpretations are available.

Categorization

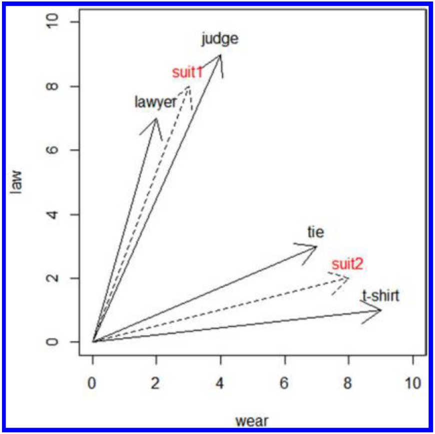
Table 2: Distribution of semantic categories in the dataset, cross-classified

	total	ABSTRACT	CONCRETE	AMBIGUOUS
EVENTIVE	275	275	0	0
NON–EVENTIVE	70	15	49	6
AMBIGUOUS	55	1	0	54
UNCLEAR	6	6	0	0
total	406	297	49	60

Vectorization

Instance vectors: one vector for each occurrence of a word.

1. The lawyer filed a **suit** to the judge.
2. The **suit** is next to the tie and the t-shirt.

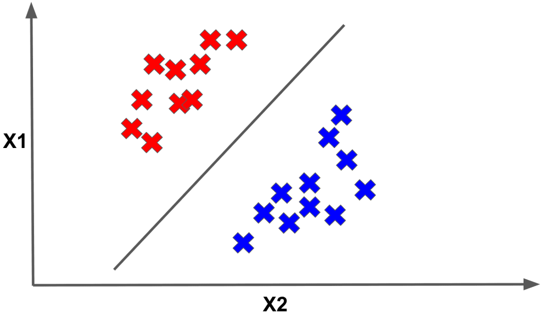


Vectorization

- Lemma vectors (Skip-gram model, context window of 5) for all context nouns, adjectives, adverbs and verbs.
- Instance vectors for target words: each target noun was assigned an instance vector made using lemma vectors of the words included in the context windows of the target noun.

Classifier

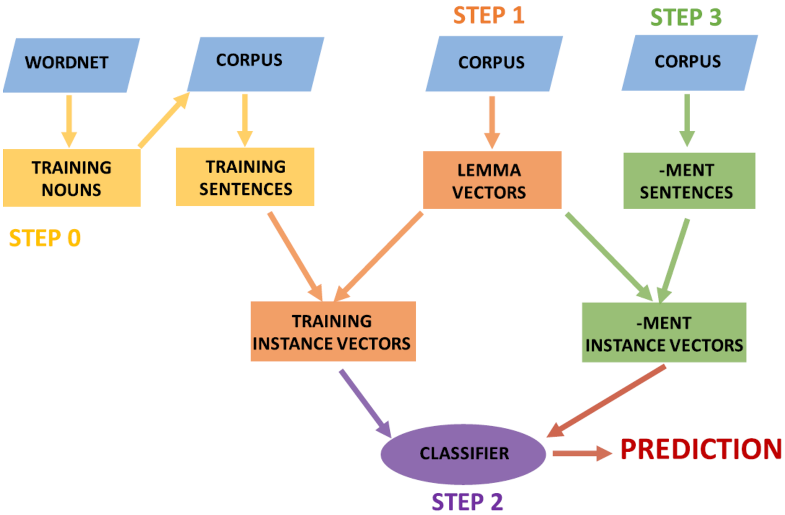
Support Vector Machine (SVM): linear separation between two sets of data to separate them into two groups.



Classification

- **Training**: frequent, unambiguous nouns from WordNet , annotated as eventive and non-eventive lax object, strict object or living thing (instance vectors).
- **Testing**: 55 ambiguous *-ment* derivatives (instance vectors).

Workflow

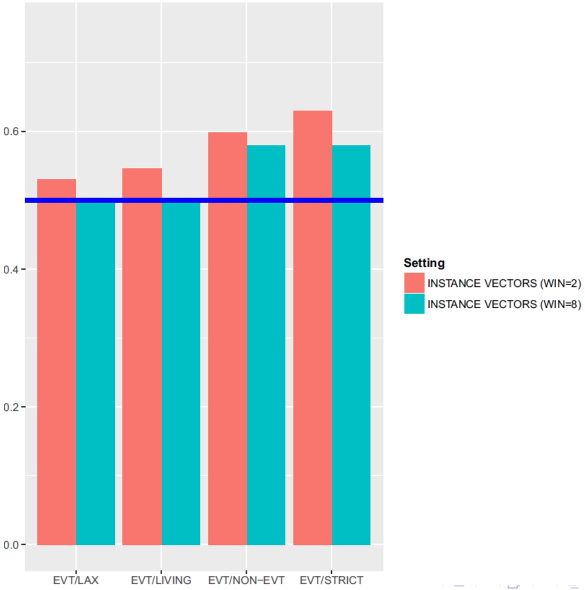

STEP 1
STEP 3
WORDNET
CORPUS
CORPUS
CORPUS
TRAINING NOUNS
TRAINING SENTENCES
LEMMA VECTORS
-MENT SENTENCES
STEP 0
TRAINING INSTANCE VECTORS
-MENT INSTANCE VECTORS
CLASSIFIER
PREDICTION
STEP 2

First experiment

Eventive / non-eventive classification of the instance vectors

- Results are not convincing $\rightarrow$ the non-eventive class is too diverse (abstract nouns, artifacts, animals...)

Main experiment



Second experiment

Table 4: Proportions of predicted interpretations vs. manual annotations (figures are rounded)

Predictions \ **Annotations**	eventive	non-eventive	ambiguous
eventive	0.78	0.47	0.73
non-eventive	0.22	0.53	0.27

Main experiment

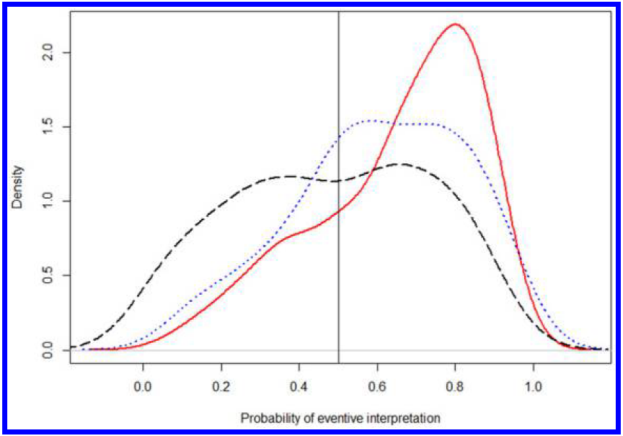


Figure 6: Distribution of predicted probabilities for *-ment* derivatives that are manually annotated as EVENTIVE (unbroken red line) vs. NON-EVENTIVE (dashed black line) vs. AMBIGUOUS (dotted blue line)

Analysis

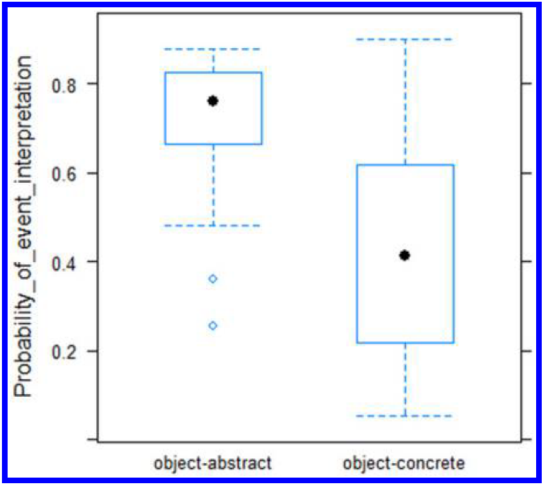


Figure 7: Distribution of predicted probabilities for concrete vs. abstract non-eventive *-ment* derivatives

Analysis

Examples of well classified eventive nouns:

- I got over that *initial moment* of **dumbfoundment** (I'm making up my own words today). (0.84)
- Hydrogen, especially atomic hydrogen, is particularly dangerous because it tends to *cause rapid* **embrittlement** even at low temperatures. (0.96)

Examples of well classified concrete nouns:

- The "U" shaped cap will cover the frame and hide the old exterior putty or *concrete* **embedment**. Once caulked, the exterior will be neat, finished, and weather tight. (0.05)
- I developed a custom fit *cardboard* **fitment** that held the USB all nice and tight with the bonus of a business card next to it for company. (0.05)

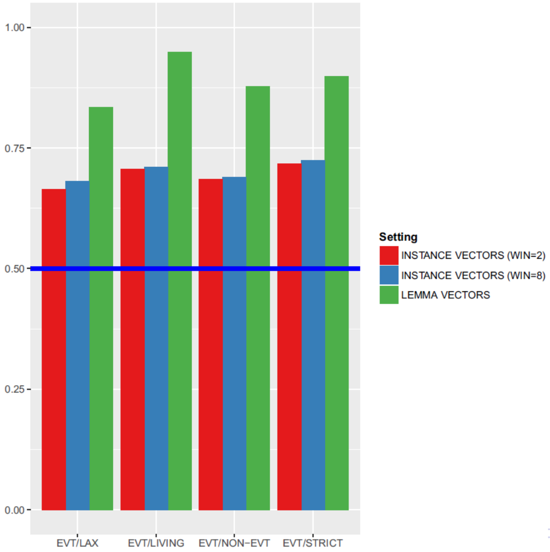
The semantic space of the **ambiguous nouns** is very similar to the semantic space of the **abstract object nouns**, which is also similar to eventive nouns.

Analysis

The semantic space of the **ambiguous nouns** is very similar to the semantic space of the **abstract object nouns**, which is also similar to eventive nouns.

Example (Melloni, 2011):
1. The **arrangement** of the book is misleading. (result state)
2. The **translation** of the book is misleading. (abstract object)

Lemma vs. Instance vectors performance on WordNet unambiguous and mid-frequent nouns



Discussion

- Context is useful for disambiguation, but most importantly, very small context windows are sufficient to make a good prediction
  - Assumption: immediate linguistic context carries the meaning because speakers do not want the listeners to mobilize too many resources when introducing a new word.
- Non-eventive derivatives are harder to predict than eventive ones
  - Ambiguous and abstract non-eventive share the same distributional properties and correlate with eventive nouns.
  - Contextual ambiguous words (i.e. nouns that fall into the abstract non-eventive and the eventive classes) are already hard to disambiguate for humans.
- Instance vectors perform better than the baseline, although not as well as lemma vectors.