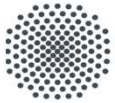


Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

Using Embeddings To Assess Cross-Lingual Frame Applicability

Jennifer Sikos

Sebastian Padó

Based on a paper presented at the 2018 COLING workshop on “Linguistic Resources for NLP”

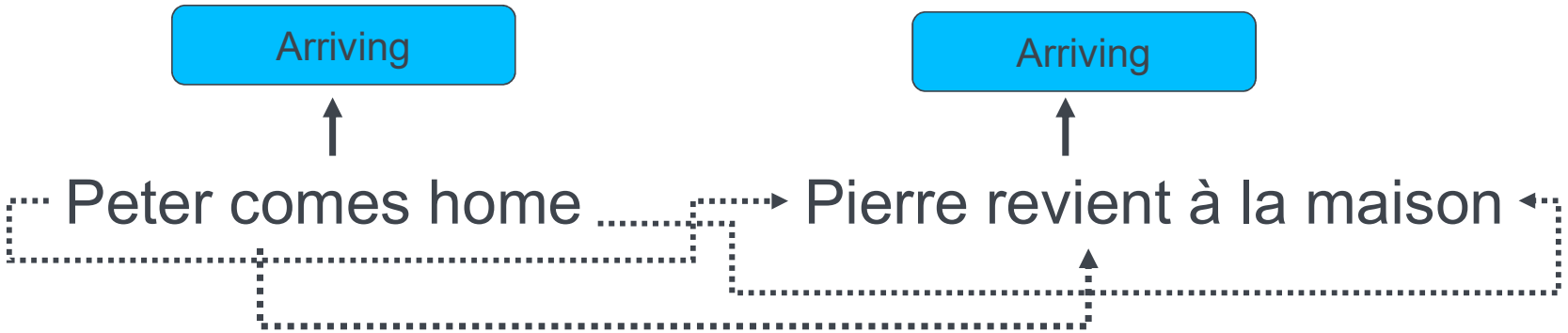
Research Question

- What can computational linguistics contribute to assessments of cross-lingual frame applicability?

- Traditionally: Linguistic/lexicographic analysis
  - Discrete
- Our proposal: Computational lexical semantics models (embeddings)
  - Continuous
  - Can we still see something interesting?

Background: Parallel Corpora

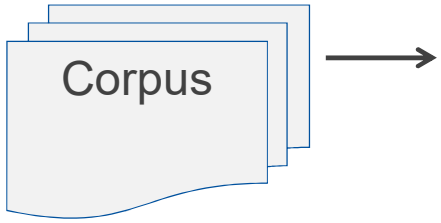
- In the past, computational methods relied on parallel corpora to align frame structures across languages
  - Frames that didn't have good alignment weren't applicable
- Annotation alignment (Erk and Padó 2006):



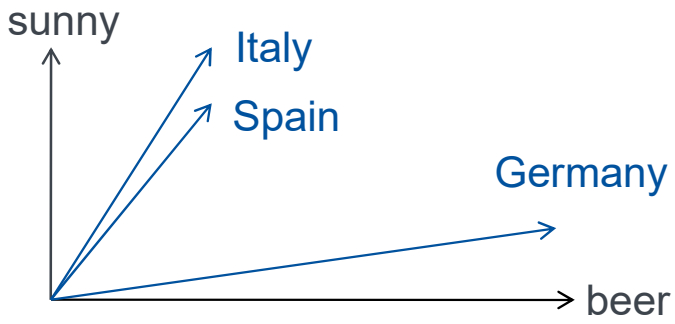
- Parallel corpora problematic for a number of reasons (e.g., scarcity)

Background: Distributional Semantics

- Instead of parallel corpora, can we aggregate over all the uses of a frame across languages and test if the uses are comparable?
- Distributional hypothesis: **Similar corpus** use of words indicates **similarity in meaning** (Firth 1957, Harris 1968)



	sunny	beer
Italy	5	1
Spain	4	1
Germany	1	20



Background: Distributional Semantics

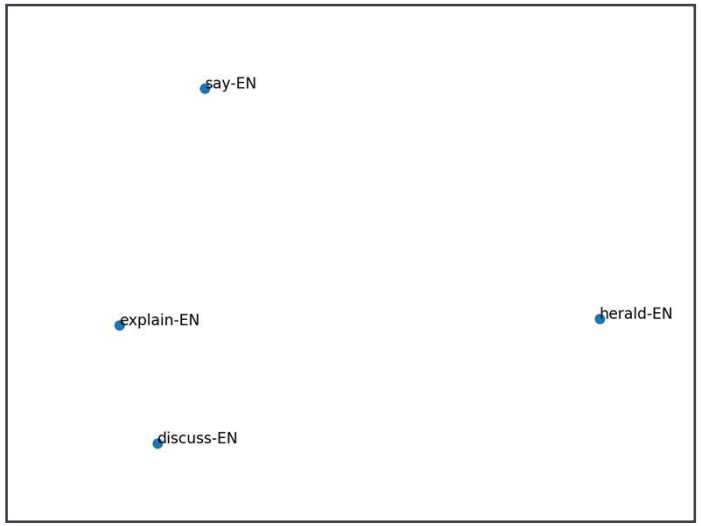
- In about 2012, Distributional Semantics was taken over by neural networks
  - Learn a network that predicts, for each word, its corpus contexts
  - Technical advantages: compact representation, good performance
  - **Embeddings**
    - Conceptually the same as count-based distributional methods

Background: Distributional Semantics

- Embeddings
  - Traditionally, based on words in a corpus
  - No reason to have such limitation – could also be based on morphemes, phrases, anything in the corpus can essentially be represented by embeddings (for example, frames)
- Method
  - **Word2Vec** (Skipgram / CBOW) / … (Mikolov et al 2013a)

Background: Embeddings

English



German

Background: Embeddings

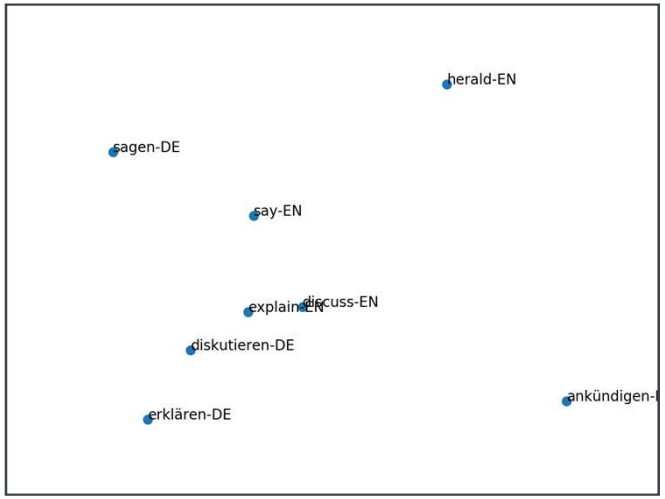
- Assuming there are seed words that share meaning across the individual spaces, we can build a mapping from one space to the other (Mikolov 2013b)
- These words can be any seed terms that share meaning, such as named entities
- By building this mapping, we can project the vectors to **a single, shared space**
- Unlike previous work, no requirement for a parallel corpus

Mapped English/German

Hypothesis

Semantic similarity in the embedding space correlates with **cross-lingual frame applicability**

Mapped English/German



Experiment: Corpora

The Washington Post ***reported*** *on the country's biological weapons lab*

Frame-Annotated Corpus: English

Lemmatized: The Washington Post report on the country 's biological weapon lab

Detect Entities: The_Washington_Post report on the country 's biological weapon lab

Annotated Frames: The Washington Post **reported** on the country's biological_weapons lab

Statement

Experiment: Corpora

*Konzernchefs **lehnen** den Milliardär als US Präsident **ab***
CEOs reject the billionaire as US President (part)

Frame-Annotated Corpus: German

Lemmatized: **Konzernchef** lehnen **der** Milliardär als US Präsident ab

Multi-word Predicates: Konzernchef **ablehnen** der Milliardär als US Präsident

Detect Entities: Konzernchef ablehnen der Milliardär als **US_Präsident**

Annotated Frames: Konzernchef **ablehnen** der Milliardär als US_Präsident

Judgment_communication

Experiment: Corpora

*The Washington Post **reported** on the country's biological weapons lab*

*Konzernchefs **lehnen** den Milliardär als US Präsident **ab***

Lexical Unit (LU) Embeddings

The_Washington_Post **report** on the country 's biological weapon lab
Konzernchef **ablehnen** der Milliardär als US_Präsident

Frame Embeddings

The_Washington_Post **STATEMENT** on the country 's biological weapons lab
Konzernchef **JUDGMENT_COMMUNICATION** der Milliardär als US_Präsident

Corpora

English FrameNet 1.5 Annotated Corpus		German SALSA Annotated Corpus	
Annotated sentences	170k	Annotated sentences	24k
LUs	11k	LUs	998
Number of annotations per LU	20.7	Number of annotations per LU	36.9

Case Study 1

Frame Specificity

- FrameNet frames exist along a hierarchy
- More specific frames are more constrained
- Frame specificity can be measured by frame relations
- Express parent/child, abstract/specific relationships: is_Inherited_by, is_Used_by, and has_Subframe
- Frames with higher counts -> more abstract

Communication

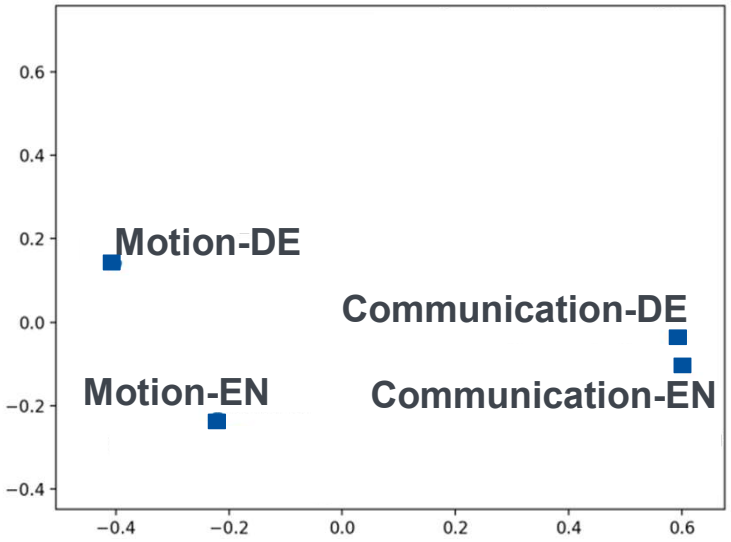
Is Inherited by

Communication_Manner

Communication_Response

Case Study 1

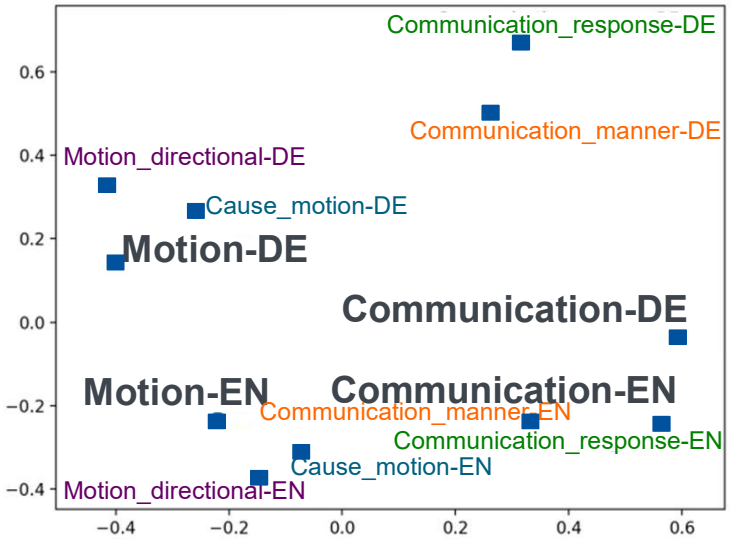
Frame Specificity



Frame (# relations)	Cross-Lingual Similarity
Communication (39)	0.54
Motion (25)	0.39

Case Study 1

Frame Specificity



Frame (# relations)	**Cross-Lingual Similarity**
Communication (39)	0.54
Communication_Manner (0)	0.17
Communication_Response (1)	0.14
Motion (25)	0.39
Cause_Motion (10)	0.15
Motion_Directional (1)	0.27

Case Study 2

Annotation choice in SALSA

- **announce**
  - *Statement* frame
  - Speaker, Message, Medium semantic roles
- **ankündigen**
  - German translation of announce that can express a Speaker, Message, and Medium in similar syntactic frames to English - not all can (Boas 2013)

They **have announced** the birth of their first child.

Sie **haben** die Geburt ihres ersten Kindes **angekündigt**.

Case Study 2

Discussion of Annotation choice in SALSA: **Ankündigen**

- Die Schatzmeister der beiden Parteien protestieren dagegen und **kündigten** juristische Schritte **an**.
  The treasurers of both parties protested and ***announced*** *legal action against it*
- Regierung in Rom **kündigt** Preisstopp und Sparprogramm **an**.
  Government in Rome ***announces*** *a spending freeze and fiscal frugality*
- Die SPD **kündigte** erneut Widerstand gegen die geplante Einführung von Karenztagen **an**.
  The SPD party once more ***announced*** *its resistance to the planned introduction of unpaid sick days*

Case Study 2

Annotation choice in SALSA

Predicate	German frames	Count
ankündigen	*Heralding*	85
	Omen	1
	Evidence	1

Heralding: A communicator announces (linguistically or non-linguistically) the imminent arrival of an individual or the occurrence of an event.

Case Study 2

Annotation choice in SALSA

Predicate	German frames	Count	Most Similar English frame
ankündigen	*Heralding*	85	***Statement***
	Omen	1	
	Evidence	1	

- Recovery of a relationship between *Heralding* and *Statement* frames
  - Possibly a Using relation? (Sikos and Padó C&F 2018)
- Measure of how a word in one language fits a frame in another

Conclusions

- Original hypothesis: semantic similarity in embedding space correlates with cross-lingual frame applicability
- Revised hypothesis: semantic similarity in embedding space correlates with comparability of annotation
  - Cross-lingual applicability *plus* monolingual annotation choices
  - Additional factors: frequency effects, coverage of lexical units, etc.
- Still a success: helps us better understand how annotations in different languages relate to one another

Future

- Deeper investigation into the relation between LUs and frame embeddings in vector space
  - LUs that are more prototypical of the frame
  - Frames as a set of LU vectors in embedding space
- Frames are comparable if there are enough LUs that correspond
- Beyond annotated data
  - Methods that don’t rely on annotations: cluster LU embeddings
  - Detect new LUs for a frame

Thank you!

Using Relation as a Source of Concept-based Paraphrases [Sikos and Padó, C&F 2018]

- Concept-based paraphrases involve common-sense knowledge

  a. [He Speaker] called [him Entity] [a hero Label].

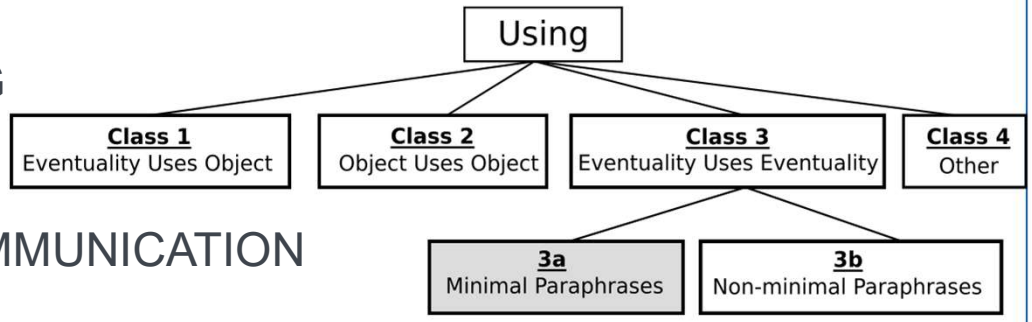
  b. [He Communicator] praised [him Evaluee] [for being a hero Reason].

- Using frame-to-frame relation: links a more concrete frame to a more abstract but related frame
  - Ex: LABELING Uses JUDGMENT_COMMUNICATION
- However, Using relation has been interpreted in different ways, leading to different frame-to-frame relations within Using

Using Relation as a Source of Concept-based Paraphrases

Examples:

- Class 1: DRESSING Uses ACCOUTREMENTS
  - "Jack put on his watch on his wrist"
- Class 2: CLOTHING_PART Uses CLOTHING
  - "Renwick bent to tie his shoelaces"
  - "Renwick bent to tie his shoes"
- Class 3: LABELING Uses JUDGMENT_COMMUNICATION
  - "He called him a hero"
  - "He praised him for being a hero"
- Class 4: TRANSLATING Uses MENTAL_ACTIVITY
  - MENTAL_ACTIVITY is not lexicalized



Using Relation as a Source of Concept-based Paraphrases

Side Conditions for Minimal Paraphrases

Condition 1: Presence of Sentiment

Ex: “He called him a hero” / “He praised him for being a hero”

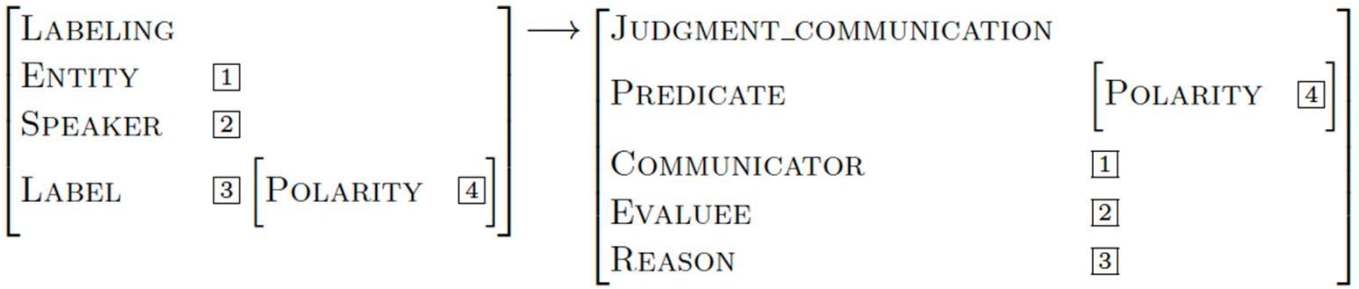


Figure 7: Side Condition 1: Presence of Sentiment.