

# An Empirical View on Semantic Roles Part V

Katrin Erk  
Sebastian Pado  
Saarland University  
ESSLI 2006

1

---

---

---

---

---

---

---

---

## Structure

1. A Historical Introduction
2. Contemporary Frameworks
3. Empirically Difficult Phenomena
4. Role Semantics vs. Formal Semantics
5. Cross-linguistic Considerations

2

---

---

---

---

---

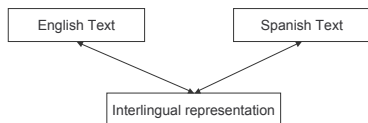
---

---

---

## The Interlingua idea

- A language-independent representation
  - Contains all relevant information (**complete**)
  - Abstracts over all language-specific phenomena (**language-independent**)
- Could be used for all kinds of cross-lingual tasks
  - Cross-lingual IR, Machine Translation...
- Completeness requires semantic information



3

---

---

---

---

---

---

---

---

## Frame Semantics as interlingua

- Is a frame-semantic analysis an interlingua?
- Short answer: no, **incomplete information**
  - Does not model (e.g.) modality, negation
  - Cf. part 4

4

---

---

---

---

---

---

---

---

## Frame Semantics as interlingua

- Cross-lingual aspects of frame semantics still interesting
  - More informative than "formal semantics" (lexical information)
    - In formal semantics, formula structure mirrors syntactic structure
  - Predicate-argument structure as *part* of interlingua
    - Lexical conceptual structure (LCS), Dorr 1990
- At least provides suitable description level to study differences (Boas 2005)
- Question: how **language-independent** are frame-semantic analyses?
  - Quick answer: To a significant degree
  - Idea of this part: Close look at cross-lingual data
  - NB: This is research territory!

5

---

---

---

---

---

---

---

---

## Language independence of frame-semantic analysis

1. Type-level appropriateness
  - Are English FrameNet frames appropriate to describe semantic classes of other languages?
2. Token-level appropriateness
  - For any pair of translated sentences  $(s_1, s_2)$ , are the frame-semantic analyses of  $s_1$  and  $s_2$  parallel?

6

---

---

---

---

---

---

---

---

## Type-level appropriateness

- Naïve assumption: FrameNet frames can be used to annotate other languages
  - Manual FrameNet-style data analysis in progress for French, German, Japanese, Spanish,...
- Works surprisingly well (for majority of frames)
  - Cited reason: "Conceptual nature of frames"
- However: for each language, some frames don't work

7

---

---

---

---

---

---

---

---

## Cross-lingual frame problems

- Review: Criteria for frame creation
  - A frame is a class of predicates that
    - Refer to the same situation and allow the same inferences about participants
    - Can realise the same set of roles
- Problems arise if languages differ in
  - Either the way they "package" situations
  - Or the way they realise arguments
- General area: Typological differences

8

---

---

---

---

---

---

---

---

## "Package" problems: Granularity of predicates

- The level of detail in semantic distinctions can vary across languages
  - English almost always distinguishes between OPERATE\_VEHICLE (as driver) and RIDE\_VEHICLE (as passenger)
    - drive: usually OPERATE\_VEHICLE (context can override)
    - ride: only RIDE\_VEHICLE
  - German does not consistently make the difference
    - fahren: subsumes both drive and ride
      - Without context: distinction not possible
      - Even **within** corpus: context often does not disambiguate
  - Right level of description for "fahren": USE\_VEHICLE
    - "Empty" (non-lexicalised) frame in English

9

---

---

---

---

---

---

---

---

## Argument realisation problems: Language-specific constructions

- German: General construction “Free dative”
  - Can realise “Affected party”
  - Constructional alternative to possessive
- Example: Frame PERCEPTION\_ACTIVE (Role Direction)
  - [auf die Koepe der Moenche<sub>DIR</sub>] schauen  
to look [onto the heads of the monks<sub>DIR</sub>]
  - [den Moenchen<sub>?</sub>] [auf die Koepe<sub>DIR</sub>] schauen  
to look [the monks<sub>?</sub>] [onto the heads<sub>DIR</sub>]
- Discontinuous role / no role / additional role?

10

---

---

---

---

---

---

---

---

## Argument realisation problems: Language-specific constructions

- Spanish motion verbs accept both PURPOSE and INTENTION frame elements
  - Voy a Malaga [para pedirle dinero a un amigo<sub>PURP</sub>]  
I'm going to Malaga [to ask a friend for Money]
  - Voy a Malaga [a ver a un amigo<sub>INT</sub>]  
I'm going to Malaga [to see a friend]
  - Voy a Malaga [a visitar a un amigo<sub>INT</sub>] [para pedirle dinero<sub>PURP</sub>]  
I'm going to Malaga [to see a friend and ask him for money].

11

---

---

---

---

---

---

---

---

## Argument realisation problems: Ontological distinctions

- In FrameNet, ontological distinctions between frame elements often complemented by language-specific syntactic characterisations
  - Example: Frame AWARENESS
    - Content: “The object of the cognizer’s awareness” -- NP/S
      - He believes [that the window is open].
    - Topic: “The subject area of the awareness” -- PPs
      - He knows [about the window]
  - Does not carry over well to German
    - Er weiss [um die Ungeduld seiner Landsleute]  
He know [about/– the impatience of his compatriots]
    - Content or Topic?

12

---

---

---

---

---

---

---

---

## [ Frames as interlingua ]

1. Type-level appropriateness
  - Are English FrameNet frames appropriate to describe semantic classes of other languages?
2. Token-level appropriateness
  - For any pair of translated sentences  $(s_1, s_2)$ , are the frame-semantic analyses of  $s_1$  and  $s_2$  parallel?

13

---

---

---

---

---

---

---

---

## [ Token-level appropriateness ]

- For any pair of translated sentences  $(s_1, s_2)$ , are the frame-semantic analyses of  $s_1$  and  $s_2$  parallel?
- Short answer: no.
  - Example 1: free translations
  - Example 2: "fahren/drive"
- We want to qualify this statement.

14

---

---

---

---

---

---

---

---

## [ Three classes of cases ]

- General picture: Three classes of predicate translations
  1. Matches (same frame)
  2. Controllable mismatches (different, but related frame)
  3. Idiosyncratic cases

15

---

---

---

---

---

---

---

---

## Parallel corpora

- Look at **word-aligned predicate pairs** in parallel corpora

- o EUROPARL

- Questions:

- o Do frames match?
  - If yes, do roles match?
  - If no, can we characterise the divergence?

Kim promised to be on time  
Kim versprach pünktlich zu kommen

```
graph TD
    K1[Kim] -.-> K2[Kim]
    P[promised] -.-> V[versprach]
    P -.-> T[to be on time]
    T -.-> P2[pünktlich zu kommen]
```

16

---

---

---

---

---

---

---

---

## Three classes of cases

- General picture: Three classes of predicate translations

1. Matches (same frame)
2. Controllable mismatches (different, but related frame)
3. Idiosyncractic cases

17

---

---

---

---

---

---

---

---

## Class 1: Perfect matches

- Corpus study to assess frequency of perfect matches:

1. Data Selection: Concentrate on "close translations"
  - o 1000 sentence pairs from English-German bitext
  - o Predicate pairs with at least one frame in common
    - read / lesen ("read") is in
    - read / herausfinden ("find out") is out
  - o FrameNet lexicon (En), SALSALSA lexicon (De)
2. Data Annotation: Give sentence pairs a frame-semantic analysis
  - o Must guarantee *independent* annotation

18

---

---

---

---

---

---

---

---

## Results

- Same frame evoked: ~72% of cases
  - Number somewhat difficult to interpret
    - Inter-annotator agreement (upper bound) was 0.85
- Good news: If same frame is evoked, 90% of roles occur in both sentences
  - Remaining differences mostly active/passive alternations:
    - En: I hope that [Ireland] will be remembered
    - De: I hope that [we] will remember [Ireland]
- For a considerable fraction of cases, the frame-semantic analysis agrees across languages
  - At least for related languages like English and German

19

---

---

---

---

---

---

---

---

## Three classes of cases

- General picture: Three classes of predicate translations
  1. Matches (same frame)
  2. **Controllable mismatches (different, but related frame)**
  3. Idiosyncratic cases

20

---

---

---

---

---

---

---

---

## Class 2: "Controllable" mismatches

- Question: Can we characterise the cases where frames do not match?
  - First look at "simple" mismatch cases
  - Study on cases where
    - we expect close semantic structure (same frames)
    - but syntax makes this impossible
  - Translation pair increase - höher (higher)
    - Details: see Pado and Erk (2005) in reader

21

---

---

---

---

---

---

---

---

## Intransitive “increase”

- **Inchoative/stative** frame: Can only realise “Item”

	<b>CHANGE_POSITION_ON_A_SCALE (CPOS)</b>
Def	This frame consists of words indicating the change of an ITEM’s position on a scale.
FES	<b>ITEM</b> The tea price rose.
FEEs	advance.v, decline.n, decline.v, decrease.n decrease.v, diminish.v, double.v, increase.v, rise.v

- Same analysis for German höher: **stative** adjective

---

---

---

---

---

---

---

---

## Example

The tea price rises.

---

---

---

---

---

---

---

---

## Transitive “increase”

- **Causative** frame: can realise both “Item” and “Cause”

	<b>CAUSE_CHANGE_OF_SCALAR_POSITION (CCOSP)</b>
Def	This frame consists of words indicating that a CAUSE affects the position of an ITEM on a scale.
FES	<b>CAUSE</b> The drought increased tea prices. <b>ITEM</b> Lipton’s increased the tea price.
FEEs	cut.n, cut.v, decrease.v, diminish.v, growth.n, increase.v, lower.v, move.v, raise.v, reduce.v

- What happens if this sense is translated with the **stative** adjective?

---

---

---

---

---

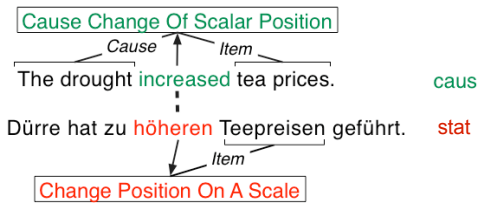
---

---

---



## [ An example ]




---

---

---

---

---

---

---

---

## [ Evaluation ]

English	German	Count
CPOS (36 n, 13 v, 24 ppart)	CPOS (adj)	73
CCOSP (49 v)	CPOS (adj)	49

- Causative/stative cases make up about 40% of all cases
  - Mismatch: No direct frame correspondence

---

---

---

---

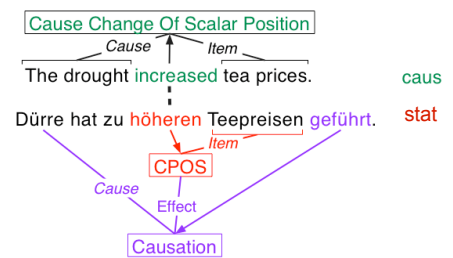
---

---

---

---

## [ What happens for causatives? ]



X increases Y == X leads to a higher Y

---

---

---

---

---

---

---

---

## Frame Group Matching Hypothesis

X increases Y == X leads to a higher Y

- Languages distribute semantic material differently among adjacent frames (**frame groups**)
- Hypothesis: If the aligned predicate pairs evoke similar frames, we can find frame groups covering **exactly** the same semantic material
  - Translation as **semantic paraphrase**

28

---

---

---

---

---

---

---

---

## Getting to frame group paraphrases

- Intuition: Identify frame groups by matching roles
- Algorithm: Start out with one known frame group
  - Iteratively identify frame groups whose roles exactly correspond to known paraphrases
    - Go back and forth between languages
    - New paraphrases

29

---

---

---

---

---

---

---

---

## Quantitative Evaluation

- 110 of 122 sentences can be explained by the paraphrase set for CCOSP
  - Group 1 (65): No Cause on either side  
An increase in X == A higher X
  - Group 2 (45): Causer on both sides  
X increases Y == X leads to a higher Y
- 12 sentences cannot be explained, due to **role mismatches**:  
X leads to a higher Y == Y increases

30

---

---

---

---

---

---

---

---

## Identified paraphrases

- CCOSP (*X increases Y*) paraphrased by CPOS plus
  - Achievement (*X achieves a higher Y*)
  - Causal\_Connection (*X is related to a higher Y*)
  - Deciding (*X decides for a higher Y*)
  - Means (*X is a means for higher Y*)
  - ...
- Related to cognitive account of causality (Talmy 2000)
  - Distinction between different "causality situations"
  - Correspond (at least partly) to our different paraphrases
    - Agentive causality  $\Leftrightarrow$  Achievement
  - Talmy's "gradience in causality": Causal\_connection

31

---

---

---

---

---

---

---

---

## Controllable mismatches: Outlook

- In our study, frame groups provide concise model for semantic variance in translations
  - Assumption: same roles realised
  - Linguistically defined handle on (simple) world knowledge
- Problem 1: "Same roles" assumption
  - Too strong in general (passives!)
- Problem 2: Validity of frame groups?
  - In the experiment, (almost) all frame groups we found were sensible
  - However, clean data and manual analysis

32

---

---

---

---

---

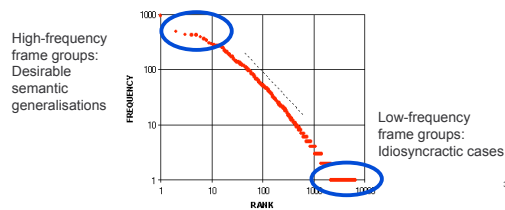
---

---

---

## Frame groups and frequency

- Large-scale automatic acquisition probably results in Zipf distribution
  - Frequency approximates validity?



33

---

---

---

---

---

---

---

---

## [ Three classes of cases ]

- General picture: Three classes of predicate translations
  1. Matches (same frame)
  2. Controllable mismatches (different, but related frame)
  3. **Idiosyncratic cases**

34

---

---

---

---

---

---

---

---

## [ Class 3: Idiosyncratic cases / Infrequent translations ]

- Question: What kinds of infrequent translations are there?
  1. Perfectly good, but infrequent translations
    - Especially problematic in specialised corpora
  2. Translations that only hold in a specific context
  3. Translation errors
  4. (Technical errors, e.g. alignment errors)

35

---

---

---

---

---

---

---

---

## [ An example ]

...questions that were not **answered** during answering time...

**Answering**

...les questions qui ne sont pas **examinées** pendant l'heure des questions...

(the questions that were not **examined** during question time)

**Scrutiny**

Frame group: Answering <-> Scrutiny

36

---

---

---

---

---

---

---

---

## “Correlated events”

- examine vs. answer
  - In the context of questions:  
A question that is examined is usually/often/mostly answered
- Other examples:
  - precaution/prevent: The purpose of a precaution is to prevent something
  - give/receive: If something is given to X, X receives it

37

---

---

---

---

---

---

---

---

## The nature of translation

- Translation is driven by conceptual considerations
  - Recreate the communicative function of the text in the target language
  - Translation can **incorporate world knowledge**
    - Linguistic form / Semantic structure may change

38

---

---

---

---

---

---

---

---

## The gradient of world knowledge

Close translation	Increasingly free translation
Semantic structures correspond	Less semantic similarity
No world knowledge	More world knowledge

- Free translations are problematic
  - Not straightforward to model
- But also a chance!
  - Bootstrapping for acquisition of world knowledge?

39

---

---

---

---

---

---

---

---

## Summary

- Frame Semantics is not an interlingua, but it has strong cross-lingual appeal
  - For a considerable number of cases, we obtain parallel analyses (class 1)
  - For a second class, we obtain analyses that are different, but in predictable ways
  - A third class comprises cases whose translation is idiosyncratic
    - Most difficult, but also most interesting

40

---

---

---

---

---

---

---

---

## Outlook

- Cross-lingual properties of FrameNet make possible automatic induction of FrameNet data for new languages
  - Idea: follow word alignments in parallel corpus to find predicates for frames and constituents for roles
- Application of frame-semantic analyses for cross-lingual information access tasks?
  - Open area for research

41

---

---

---

---

---

---

---

---

## References

- H. Boas: Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography* 18(4), 2005.
- Burchardt, Erk, Frank, Kowalski, Pado, and Pinkal: The SALSA Corpus: A German corpus resources for lexical semantics. *Proceedings of LREC 2006*.
- S. Pado and K. Erk: To cause or not to cause: Cross-lingual semantic matching for paraphrase modelling. *Proceedings of the Cross-Language Knowledge Induction Workshop 2005*.
- S. Pado and M. Lapata: Cross-lingual projection of role-semantic information. *Proceedings of HLT/EMNLP 2005*.
- C. Subirats and H. Sato: Spanish FrameNet and FrameSQL. *Proceedings of LREC 2004*.
- L. Talmy: *Towards a Cognitive Semantics*, chapter The Semantics of Causation. MIT Press, 2000.

42

---

---

---

---

---

---

---

---

## References - FrameNets for other languages

- SALSA (German FrameNet)  
<http://www.coli.uni-saarland.de/projects/salsa/>
- Spanish FrameNet  
<http://gemini.uab.es/>
- Japanese FrameNet  
<http://jfn.st.hc.jkeio.ac.jp/>

---

---

---

---

---

---

---

---