An Empirical View on Semantic Roles
Part V

Katrin Erk
Sebastian Pado
Saarland University
ESSLLI 2006

Structure

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

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

[Diagram of Interlingua representation with English Text, Spanish Text, and Interlingual representation]
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

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
- LCS (Lexical conceptual structure), 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!

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?
Type-level appropriateness

Naive 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

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

"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
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 Köpfe der Monche] schauen to look [onto the heads of the monks]
  - [den Monchen] [auf die Köpfe] schauen to look [the monks] [onto the heads]
- Discontinuous role / no role / additional role?

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] I’m going to Malaga [to ask a friend for money]
  - Voy a Malaga [a ver a un amigo] I’m going to Malaga [to see a friend]
  - Voy a Malaga [a visitar a un amigo] [para pedirle dinero] I’m going to Malaga [to see a friend and ask him for money].

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" — NPs
    - 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?
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?

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.

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
Parallel corpora

- Look at word-aligned predicate pairs in parallel corpora
  - EUROPARL
- Questions:
  - Do frames match?
  - If yes, do roles match?
  - If no, can we characterise the divergence?

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

Class 1: Perfect matches

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

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

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
**Intransitive “increase”**

- **Inchoative/Statative frame: Can only realise “Item”**
  
  **Definition:** This frame consists of words indicating the change of an ITEM’s position on a scale.

<table>
<thead>
<tr>
<th>ITEM: The tea price rose.</th>
</tr>
</thead>
</table>

| Verb: advance v, decline n, decline v, decrease n, decrease v, diminish v, double v, increase n, rise v |

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

---

**Example**

The tea price rises.

---

**Transitive “increase”**

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

  **Definition:** This frame consists of words indicating a CAUSE that affects the position of an ITEM on a scale.

  | Cause: The drought increased tea prices. |
  | ITEM: Rupert’s increased the tea price. |

- **Verb:** cut, cut v, decrease v, diminish v, grow up, increase v, lower v, move v, raise v, reduce v

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

<table>
<thead>
<tr>
<th>Cause Change Of Scalar Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause</td>
</tr>
<tr>
<td>Item</td>
</tr>
<tr>
<td>The drought increased tea prices. caus</td>
</tr>
<tr>
<td>Düre hat zu höheren Teepreisen geführt. stat</td>
</tr>
<tr>
<td>Change Position On A Scale</td>
</tr>
</tbody>
</table>

Evaluation

<table>
<thead>
<tr>
<th>English</th>
<th>German</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPOS (36 n, 13 v, 24 ppart)</td>
<td>CPOS (adj)</td>
<td>73</td>
</tr>
<tr>
<td>CCOSP (49 v)</td>
<td>CPOS (adj)</td>
<td>49</td>
</tr>
</tbody>
</table>

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

What happens for causatives?

<table>
<thead>
<tr>
<th>Cause Change Of Scalar Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause</td>
</tr>
<tr>
<td>Item</td>
</tr>
<tr>
<td>The drought increased tea prices. caus</td>
</tr>
<tr>
<td>Düre hat zu höheren Teepreisen geführt. stat</td>
</tr>
<tr>
<td>Cause</td>
</tr>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>CPOIS (adj)</td>
</tr>
<tr>
<td>Caution</td>
</tr>
</tbody>
</table>

X increases Y == X leads to a higher Y
Frame Group Matching

Hypothesis

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

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

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 leads to a higher X
  - Group 2 (45): Causer on both sides
    - X increases Y leads to a higher Y

- 12 sentences cannot be explained, due to role mismatches:
  - X leads to a higher Y == Y increases
Identified paraphrases

- CCOSP \( (X \text{ increases } Y) \) paraphrased by CPOS plus
  - Achievement \( (X \text{ achieves a higher } Y) \)
  - Causal\_Connection \( (X \text{ is related to a higher } Y) \)
  - Deciding \( (X \text{ decides for a higher } Y) \)
  - Means \( (X \text{ is a means for higher } Y) \)
  - ...
- 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

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

Frame groups and frequency

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

---

![Graph showing frequency distribution of frame groups]
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

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)

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
**“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

**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

**The gradient of world knowledge**

- Close translation: Semantically equivalent
- Increasingly free translation: 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?
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

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

References

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/