Introduction to FrameNet

Johannes Dellert

Seminar für Sprachwissenschaft, Universität Tübingen

December 18, 2007

1) Introduction

What is FrameNet?

- an electronic lexical resource based on semantic frames
- some people are inclined to call it an ontology
- based on contemporary English, mostly on the British National Corpus (BNC)
- goal: easy access for both computers and humans
- a very work-intensive project that involves a lot of manual annotation
- VERY influential, it has spawned a vast amount of literature

Who is behind this?

- the project is housed at the International Computer Science Institute in Berkeley, California
- Webpages: http://framenet.icsi.berkeley.edu/
- founder and supervisor: Charles J. Fillmore
  - Emeritus Professor of Linguistics at the University of California
  - one of the founders of cognitive linguistics
  - developed the theories of Case Grammar and Frame Semantics

Coverage and size of the database

- 10,000 lexical units grouped into 800 frames
- 120,000 example sentences taken from corpora
- definitions from the Oxford Dictionary
- links to WordNet

World knowledge in NLP

- computers as **symbol-processing devices** can use syntactic structures to apply computational semantics
- as a result, we get formal descriptions of the content of sentences: $\forall x : \mathit{human}(x) \rightarrow \mathit{mortal}(x)$

World knowledge in NLP

- computers as **symbol-processing devices** can use syntactic structures to apply computational semantics
- as a result, we get formal descriptions of the content of sentences: $\forall x : human(x) \rightarrow mortal(x)$
- however, the information transferred by a sentence is by no means always explicitly encoded
- when humans interpret and understand messages, they always take non-linguistic background information into account
- a good part of this information is contained in innumerably many little facts that humans learn during their lives and that can be subsumed as **world knowledge**
- simple example: "I went to Berlin last year." - "Had you been to Germany before?"
- for computers to "understand" those implications, we need a way to encode deep semantics

Example

"Lucy was a widow for three years."

- for "widow", in a standard lexicon as used for NLP we might find entries like
  - widow N[num=sg,gen=f]
    widow V[subcat={akk}]
- How is an IR system supposed to infer
  - that Lucy was married once?
  - that her husband died?
  - that she either died or married again after three years?

The FrameNet project tries to address some of these issues.

2) Background

A quotation

Fillmore et al. 2001

"The basic assumption of **Frame Semantics** as it applies to the description of lexical meanings is that each **word** (in a given meaning) **evokes** a particular **frame** and possibly **profiles** some **element** or aspect of that frame."

- Frame Semantics is not inherently compositional
- in Construction Grammar, also constructions contribute to the meaning

What is a semantic frame?

- a schematic representation of a situation type
- contains fields for participants, props, contextual roles
- these are called Frame Elements (FEs)
- Lexical Units (LUs) are associated and grouped into frames

An example: Revenge

LUs: “revenge” “avenge” “retaliation” etc.

Frame Elements:

Agent - the avenging person

Offender - original offender and victim of the retaliation

Injured_Party - victim of the original offense

Injury - the original offense

Punishment - the measure taken by the avenger

An example: Revenge

FEs: Agent, Offender, Injured_Party, Injury, Punishment

Example sentences:

- [Sam's brothers $_{Agent}$] AVENGED [him $_{Injured\_Party}$].
- The team sought REVENGE for their 4-1 defeat last month.
- Jo had the affair as a kind of REVENGE against Pat.
- The team took REVENGE with a resounding victory.
- The military decapitated several prisoners to AVENGE the blood of the fallen soldiers.

Granularity

- **Problem:** There are subtle differences between LUs within one frame (as seen with "retaliate" vs. "avenge")
- in fact, given the assumption that there is no true synonymy, the context and its meaning should vary between all LUs
- this is taken to be mostly a matter of granularity in the analysis
- for most purposes, not capturing all the nuances will suffice
  $\longrightarrow$ "better slightly imprecise deep semantics than none at all"

Profiling

Fillmore et al. 2001

"[...] a profiled entity is the component of a frame that integrates directly into the semantic structure of the surrounding text or sentence."

- those FEs that are *prominently contributing* to the interpretation of a sentence are called **profiled**
- example: "undergo an operation" profiles the patient, "perform an operation" the surgeon
- it is also possible that an entire frame is profiled by one position in a sentence:
  "[Sylvia's replacement by Adam on the committee] was a surprise to everyone."

Frame Elements vs. Thematic Roles

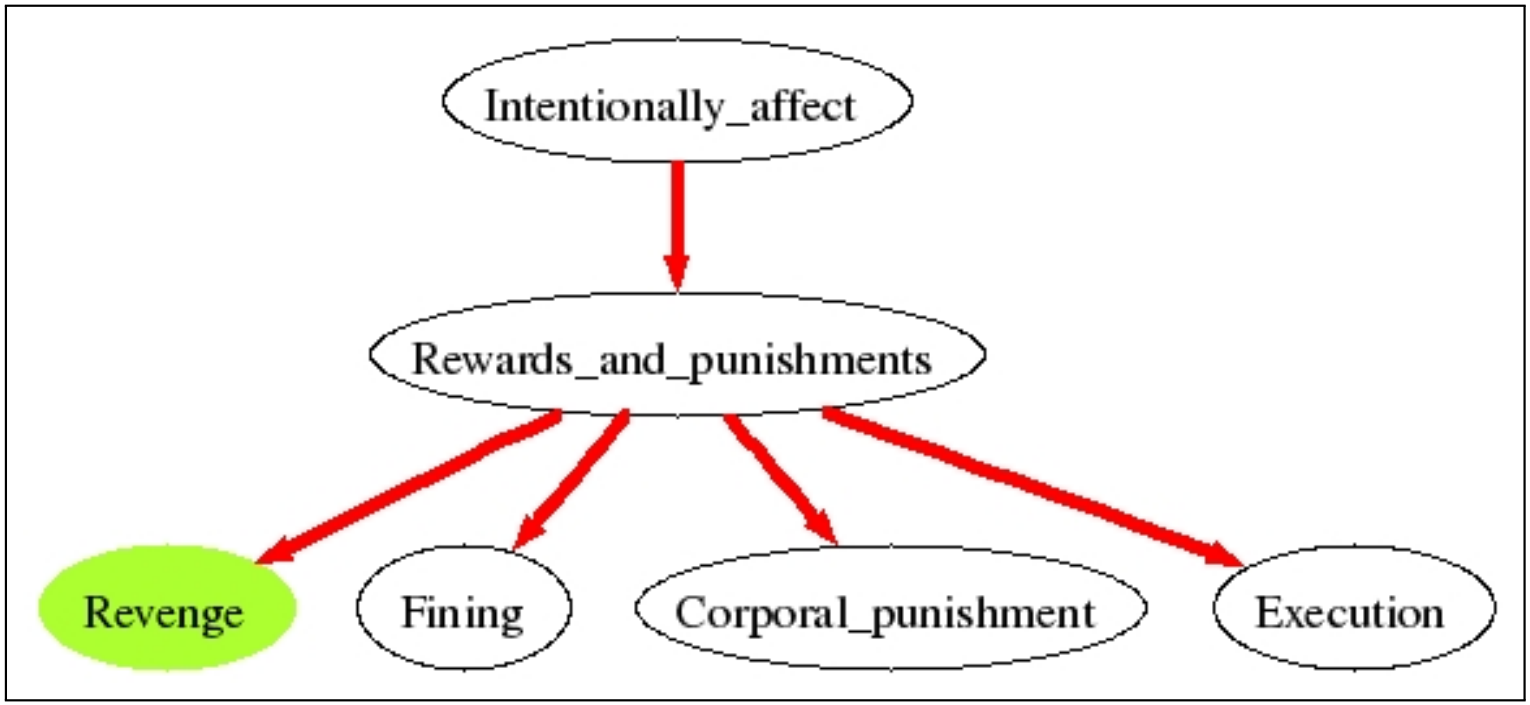
- FEs are in many ways analogous to **thematic roles** in other frameworks
- however, FEs are conceptually local to their frames
  $\longrightarrow$ no need to rely on a restricted a priori set of elements
- still, the FE labels are designed to be reusable and efforts are and will be made to keep similar roles identically labeled across frames
- however, there is no commitment to avoid adding roles if there is no obvious fit
- work can proceed without deciding on an initial complete inventory of assigned roles

Frame Inheritance

- two frames are in an inhertiance relation if one of them has all of the properties of the other plus something else
- simple example: abstract frames bequeath FEs to frames with specific kinds of profiling
  Caused_Motion is a parent frame for Insert (GOAL-profiled) and remove (SOURCE-profiled)
- there is also multiple inheritance: an execution is both a punishment and a killing

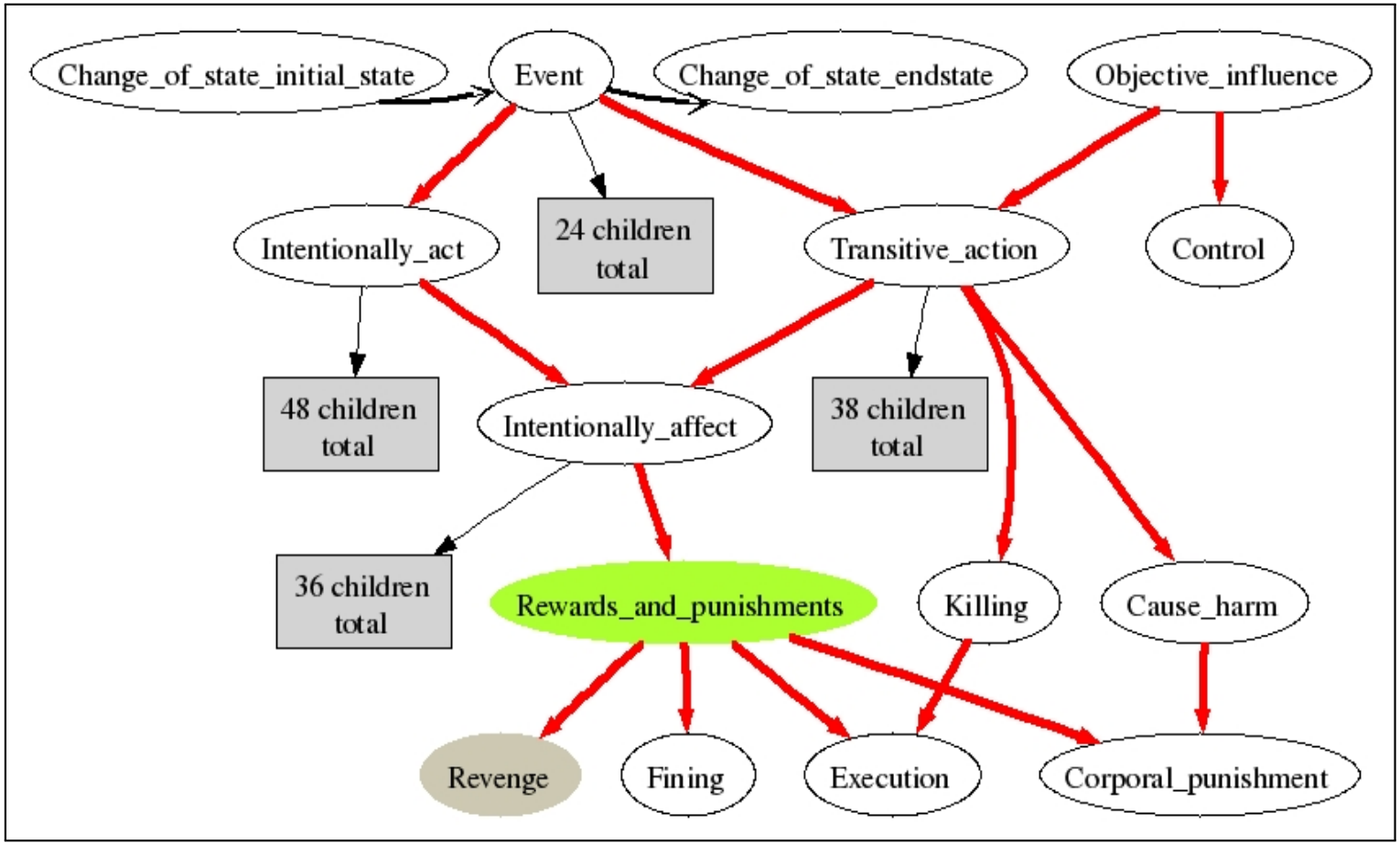
Frame Hierarchy

- given these relations, it is possible to build hierarchical structures between frames that resemble ontologies
- this is the closest vicinity of the Revenge frame:



Frame Hierarchy

- since we have multiple inheritance, the resulting structure is not a tree, but still a directed graph
- this gives an impression of the Revenge frame's larger vicinity:



Frame Composition

- when describing more complex situations, it is sometimes helpful to embody existing subframes
- to this end, a "used by"-relation is imposed on the frame set
- Example:
  - assume a frame *Own* with the FEs OWNER and POSSESSION
  - now define a new frame *Exchange* by means of *Own*:
  - *Exchange* will use two pairs of *Own* subframes with interchanged POSSESSIONs
  - note that there is a temporal relation between the two *Own* pairs, mirroring the way the two *Own* frames have changed during the exchange
  - *Exchange* therefore constitutes a simple example of a **scene**

3) Annotation

Preparation

- the native speaker analyst intuitively judges that some particular conceptual pattern underlies one or more lexical units in the language in a systematic way
- in a thorough dicussion, the structure of the frame is worked out and extended to cover the desired situations
- an initial description of the frame with its FEs is devised
- additional target words as major LUs to be associated with the frame are chosen and evaluated

Subcorpus Extraction

- sample sentences are automatically extracted from the corpora by various tools
- using syntactic patterns and a cascade filter representing a partial regular-expression grammar over POS-sequences of the language, only potentially relevant sentences are chosen
- sample sentences are boiled down to an appropriate number
- if the desired syntactic patterns couldn't be found, examples are chosen manually
- BUT: made-up sentences are a taboo

Annotation

- annotators mark selected constituents in the extracted subcorpora according to the frame elements they realize
- identification of canonical examples
- identification of novel patterns that were not yet present in the frame draft and are potential candidates for extensions
- identification of problem sentences that might entail a need for further discussion

Entry Writing

- review of the skeletal lexical entries created during the preparation phase
- inspection of the annotated sample sentences
- writing the final entries for the lexical database and the frame database
- as a result, a lexical database entry consists of:
  - the headword
  - the background frame assigned to this meaning of the word
  - a definition from the Concise Oxford Dictionary
  - the table of frame element realizations
  - the table of valence patterns
  - several annotated example sentences

4) Usage in NLP

Word Sense Disambiguation

- syntactic and collocational information is often uniquely associated with a single sense
- these patterns can be used to infer in which sense a word is used given some syntactic and collocational information
- example: disambiguating "to argue"
  - the "quarrel" sense can reliably be selected in the neighborhood of "with"-phrases or "over"-phrases
  - the "reasoning" sense is selected in th epresence of "that"-clauses and PPs with "for" or "against"
- such information is encoded in the entries for LUs in framenet
- frequency information will be added as soon as automatic FE labeling is available

Machine Translation

- FrameNets for different languages can be mapped to each other, which is currently done in pilot projects
- those mappings reveal cross-linguistic differences in the meanings and grammatical behavior of words belonging to shared frames
- larger FrameNet projects also exist for German, Spanish and Japanese, smaller ones for a variety of other languages
- vocabulary coverage will be increased in a number of restricted areas

Information Extraction

- frames provide a high precision seed for building information extraction patterns
- information extraction with semantics only based on statistical measures is very error-prone
- frames could be used as a stable backbone for matching and evaluating patterns

Question Answering

- frames facilitate the recognition of semantic relations between queries and indexed documents
- this helps to go beyond the "bag of words" approach when determining document relevance
- especially for languages with highly ambiguous lexical entries like English or Chinese, this could mean a much better recognition of the user's intentions

References

- Collin F. Baker et al. (1998)
  The Berkeley FrameNet Project
- Charles J. Fillmore et al. (2001)
  Building a Large Lexical Databank Which Provides Deep Semantics
- Collin F. Baker and Josef Ruppenhofer (2002)
  FrameNet's Frames vs. Levin's Verb Classes
- Josef Ruppenhofer et al. (2006)
  FrameNet II: Extended Theory and Practice

Thank you.