Frame Semantics and FrameNet

Miriam R. L. Petruck
miriamp@icsi.berkeley.edu
Road Map

• Frame Semantics and FrameNet
  ◦ Frames, Frame Elements, Lexical Units
  ◦ Annotation and Reports
  ◦ Frame Relations and Frame Element Relations

• Construction Grammar and Constructicons
  ◦ (Berkeley) Construction Grammar
  ◦ The FrameNet Constructicon
  ◦ Constructicon Development
What is Frame Semantics?

• an approach to the understanding and description of the meanings of lexical items (and grammatical constructions) which assumes that to understand the meanings of the words in a language we must first have knowledge of the semantic frames
What is a Semantic Frame?

- a schematic representation of a situation, object, event, or relation providing the background structure against which words are understood
Semantic Frames: Examples

- Situation: Being_attached, Being_necessary
- Event: Communication, Hiring, Replacing
- Object: Buildings, Containers, Intoxicants
- Relations: Locative_relation
What is FrameNet?

- a computational lexicography project based on the principles of *Frame Semantics*
- building a lexical resource for contemporary English that provides a body of semantically and syntactically annotated sentences from which reliable information can be reported on the *valences*, or combinatorial possibilities, of each item analyzed.
- using corpus evidence to document findings
FrameNet

- the most sophisticated instantiation of Frame Semantics

- http://framenet.icsi.berkeley.edu
FrameNet: Current Status

- 1,224 frames
- 13,640 lexical units
- 10,542 frame elements
- 1,876 frame-to-frame relations
- 202,229 annotated sentences
  - 14% “full-text” annotation
Avoiding

An Agent avoids an Undesirable_situation under certain Circumstances, where that situation may be an event or an activity.
Avoid. v

- We must avoid jumping to conclusions.
- Victoria avoids weekend work.
- They could not avoid criticism.
- Young people can avoid getting into trouble.
A FrameNet Frame

- schematic representation of a situation involving various participants, props, and other conceptual roles, each of which is a frame element
- provides the background and motivation for the existence and use of a lexical item (broadly defined) in a language
What is a Frame Element?

• **frame-specific** semantic role, the linguistic realization of which highlights different participants (or semantic roles) of a frame
Core Frame Elements (FEs)

- **Agent**: the person who avoids the **Undesirable_situation**

- **Undesirable_situation**: the situation that the Agent avoids
Annotations

[We Agent] must avoid [jumping to conclusions Undesirable_situation].

[Victoria Agent] avoids [weekend work Undesirable_situation].

[They Agent] could not avoid [criticism Undesirable_situation].

[Young people Agent] can avoid [getting into trouble Undesirable_situation].
Non-core Frame Elements

Circumstances: the Circumstances under which the Undesirable_situation is avoided

Manner: the Manner in which the Agent avoids the Undesirable_situation

Means: the Means by which an Agent avoids the Undesirable_situation
Avoiding: avoid.v - VPing/Dep

[He Agent] avoided [giving her a direct answer Undesirable_situation] [when she asked Circumstances].

[The Art Loss Register Agent] [carefully Manner] avoids [probing into the subject Undesirable_situation].

[By being subtle Means]...[you Agent] avoid [appearing naïve Undesirable_situation].
Frame Element

- Semantic role
  - Agent, Undesirable\_situation

- Grammatical Function
  - External, Object, Dependent

- Phrase Type
  - NP, VPing, PP(of various types), etc.
What is a Lexical Unit (LU)?

• Pairing between a lemma and a frame
  ◦ *hot* - It’s hot outside today.
    • hot – ambient temperature
  ◦ *hot* - The curry is really hot.
    • hot – spiciness
  ◦ *hot* - She’s one hot lady.
    • hot – desireability
Road Map

- Frame Semantics and FrameNet
  - Frames, Frame Elements, Lexical Units
  - Annotation and Reports
  - Frame Relations and Frame Element Relations

- Construction Grammar and Constructicons
  - (Berkeley) Construction Grammar
  - The FrameNet Constructicon
  - Constructicon Development
Reports

- Annotation Report
- Lexical Unit Report
Annotation Report

- 400-s20-quote
  1. Keep your hair off your face, and AVOID flat * styles."
  2. He added that we wanted to AVOID the alarming imbalance in some schemes [a reference to the German method — see above] between what is collected and the ability to reprocess it.
- 410-s20-annotate
  1. At the Congress of Berlin, the "honest broker": AVOID a Balkan conflict, but only. Russians thought, at the cost of letting down his friends.
- 429-s20-comment
  1. It is becoming harder to AVOID the conclusion that the government made an error of judgment when it decided to take Britain into the ERM in October 1990, just as Germany was entering an economic crisis.
  2. "We must AVOID running to conclusions," replied Thaeodoric cautiously.
  3. This perspective AVOIDS both general statements and general conclusions about the mass media and favours an approach which highlights the richness of the relationship between the mass media and politics.
  4. After discussing his "poor" problem, try to AVOID the weak conclusion that "it is perhaps futile."
- 429-s20-comment
  1. The "good loser syndrome" has to be AVOID intelligently.
  2. This is less odd than it looks: it pays to AVOID the delays and legal costs of chapter 3.
  3. A greater degree of regional self-sufficiency in many basic food-stuffs, especially milk and meat, could be encouraged in order to AVOID the mounting costs of transporting such commodities over long distances which is now common.
  4. It may be quicker for someone who hears well to reply, particularly when they know the answer and the question is unimportant, but this must be AVOIDED at all costs.
  5. It is hoped this will attract customers wanting to AVOID the costs and anxiety of development.
  6. And anyway, there was a current in this house, something which is not available except to AVOID at all costs.
- 429-s20-comment
  1. They could not AVOID criticism but they could usually ignore it.
  2. AVOID interpreting if you've made a criticism in good faith.
  3. By controlling me they believed that they were AVOIDING my criticism of them as parents.
  4. It is on the basis of these distinctions that they seek to AVOID certain criticisms that have been applied to Oakechott.
  5. This approach, whilst employing a similar set of areas of experience to that employed in Curriculum 11-16 and the Munn Report, AVOIDS the criticisms of ignoring teaching methods.
  6. The kindness in his voice AVOIDED suspicion and disillusion.
  7. "Let me," in a moment her quick fingers had Nicandra's gardenia and its drift of green spray firmly in place.
  8. This is to AVOID the type of criticism which may have repercussions on the team member's status or salary.
- 429-s20-comment
  1. Nobody's listening but I insist: This fairness of skin obliges them to shield themselves so that if they do catch the sun they AVOID the obvious dangers of peeling.
  2. Their old argument was this: by unilaterally renouncing nuclear weapons the West could somehow AVOID the dangers of the nuclear age.
  3. At low doses of omeprazole doctors believe they can AVOID any dangerous effects.
  4. Those who opposed eradication were much better able to survive and AVOID danger, to find food and to reproduce, multiplying ever more rapidly.
  5. The Russians could AVOID the danger of encirclement by withdrawing entirely from the salient, or advance in overwhelming numbers to seize both East Prussia and Galicia.
  6. It enables them to AVOID committing suicide and to retreat in front of danger, and this is one of the major reasons why people speak indirectly.
  7. But unless the plaintiff in unpopular or unpleasant, he certainly will normally be advised to AVOID harassment, because of the danger of heavier damages.
  8. And if you explain to Thomas that he's not to touch things like that, but must concur and tell you when he finds one, then you'll AVOID any danger in future.
  9. There were three exercises including answering questions on crime prevention, studying a video and devising a board game to illustrate how young people can AVOID getting into trouble or danger.
## Lexical Unit Report: Table 1

### Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

<table>
<thead>
<tr>
<th>Frame Element</th>
<th>Number Annotated</th>
<th>Realization(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>(145)</td>
<td>CNI:-- (35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NP.Ext (107)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[by].Dep (3)</td>
</tr>
<tr>
<td>Circumstances</td>
<td>(12)</td>
<td>PP[at].Dep (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[by].Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[without].Dep (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sub.Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[as].Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[ing][in].Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sfin.Depl (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sinterrog.Dep (2)</td>
</tr>
<tr>
<td>Degree</td>
<td>(9)</td>
<td>AVP.Dep (9)</td>
</tr>
<tr>
<td>Explanation</td>
<td>(3)</td>
<td>PP[for].Dep (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[out].Dep (1)</td>
</tr>
<tr>
<td>Manner</td>
<td>(12)</td>
<td>AVP.Dep (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[like].Dep (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[without].Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QUO.Dep (1)</td>
</tr>
<tr>
<td>Means</td>
<td>(18)</td>
<td>PP[by].Dep (13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[by].Dep (5)</td>
</tr>
<tr>
<td>Place</td>
<td>(11)</td>
<td>PP[around].Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[in].Dep (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AVP.Dep (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PP[at].Dep (4)</td>
</tr>
</tbody>
</table>
### Valence Patterns:

These frame elements occur in the following syntactic patterns:

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Agent</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 TOTAL</td>
<td>Agent</td>
<td>Manner</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>AVP Dep</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>VPing Dep</td>
</tr>
<tr>
<td>3 TOTAL</td>
<td>Agent</td>
<td>Undesirable situation</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>NP Obj</td>
</tr>
<tr>
<td>1 TOTAL</td>
<td>Agent</td>
<td>Degree</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>AVP Dep</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>NP Obj</td>
</tr>
<tr>
<td>10 TOTAL</td>
<td>Agent</td>
<td>Undesirable situation</td>
</tr>
<tr>
<td></td>
<td>CNI</td>
<td>NP Ext</td>
</tr>
<tr>
<td></td>
<td>CNI</td>
<td>NP Ext</td>
</tr>
<tr>
<td></td>
<td>CNI</td>
<td>NP Ext</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>NP Obj</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>NP Obj</td>
</tr>
<tr>
<td></td>
<td>NP Ext</td>
<td>NP Obj</td>
</tr>
</tbody>
</table>
Valence

- semantico-syntactic combinatorial possibilities
  - meaning-form-function mappings
  - FrameNet Valence Description
    - Frame Element
    - Grammatical Function
    - Phrase Type
## Avoiding.valence

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>(140) TOTAL</td>
<td>Agent</td>
</tr>
<tr>
<td></td>
<td>Undesirable situation</td>
</tr>
<tr>
<td>(15)</td>
<td>CNI</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
<tr>
<td>(17)</td>
<td>CNI</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Obj</td>
</tr>
<tr>
<td>(2)</td>
<td>CNI</td>
</tr>
<tr>
<td></td>
<td>VP[ing]</td>
</tr>
<tr>
<td></td>
<td>Dep</td>
</tr>
<tr>
<td>(89)</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Obj</td>
</tr>
<tr>
<td>(1)</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
<tr>
<td></td>
<td>PP[including]</td>
</tr>
<tr>
<td></td>
<td>Dep</td>
</tr>
<tr>
<td>(3)</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
<tr>
<td></td>
<td>Sing</td>
</tr>
<tr>
<td></td>
<td>Dep</td>
</tr>
<tr>
<td>(9)</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
<tr>
<td></td>
<td>VP[ing]</td>
</tr>
<tr>
<td></td>
<td>Dep</td>
</tr>
<tr>
<td>(1)</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Obj</td>
</tr>
<tr>
<td></td>
<td>VP[ing]</td>
</tr>
<tr>
<td></td>
<td>Dep</td>
</tr>
<tr>
<td>(3)</td>
<td>PP[by]</td>
</tr>
<tr>
<td></td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
</tr>
</tbody>
</table>
Road Map

- Frame Semantics and FrameNet
  - Frames, Frame Elements, Lexical Units
  - Annotation and Reports
  - Frame Relations and Frame Element Relations

- Construction Grammar and Constructicons
  - (Berkeley) Construction Grammar
  - The FrameNet Constructicon
  - Constructicon Development
Frame-to-Frame Relations in FN

- Inheritance
- Using
- Subframes
- Precedes
- Perspective_on
- See also
- Inchoative_of
- Causative_of

regular lexical relations
# Frame-to-Frame Relations in FN

<table>
<thead>
<tr>
<th>Relation</th>
<th>Super_frame</th>
<th>Sub_frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inheritance</td>
<td>Parent</td>
<td>Child</td>
</tr>
<tr>
<td>Subframes</td>
<td>Complex</td>
<td>Component</td>
</tr>
<tr>
<td>Precedes</td>
<td>Earlier</td>
<td>Later</td>
</tr>
<tr>
<td>Using</td>
<td>Parent</td>
<td>Child</td>
</tr>
<tr>
<td>Perspective_on</td>
<td>Neutral</td>
<td>Perspectivized</td>
</tr>
<tr>
<td>See_also</td>
<td>Main Entry</td>
<td>Referring Entry</td>
</tr>
<tr>
<td>Inchoative_of</td>
<td>Inchoative</td>
<td>State</td>
</tr>
<tr>
<td>Causative_of</td>
<td>Causative</td>
<td>Inchoative/State</td>
</tr>
</tbody>
</table>
• relationship between a more general frame, the parent frame, and a more specific one, the child
• Child frame elaborates parent frame
• Corresponding entities, FE, frame relation, and semantic characteristics, in both child and parent
• Entity is the same or more specific in child frame
• child frame = a "kind-of" parent frame

Arriving is a kind-of Motion
Inheritance

Commerce_buy inherits from Getting

Pre_getting ➔ Getting ➔ Post_getting

Inheritance ➔ Commerce_buy
Paraphrase: Inheritance

BUYING inherits from GETTING.

- Example: *I BOUGHT a new telephone*
- Paraphrase: *I GOT a new telephone.*

- paraphrase structure parallels original
Employment Scenario
Road Map

- **Frame Semantics and FrameNet**
  - Frames, Frame Elements, Lexical Units
  - Annotation and Reports
  - Frame Relations and Frame Element Relations

- **Construction Grammar and Constructicons**
  - *(Berkeley)* Construction Grammar
  - The FrameNet Constructicon
  - Constructicon Development
A **lexicon** should specify the grammatical affordances of its entries; a **grammar** should specify the kinds of lexical units capable of occurring in specifiable positions within grammatical constructions. The most consistent way to represent such mutual dependencies would be to provide both kinds of information in a single well-articulated **grammar + lexicon** (2006: 35).
Construction Grammar

...holds the view that constructions are the basic building blocks of (a) language
Construction Grammar

- Construction: a pairing of form (i.e., any combination of syntactic, morphological, or prosodic features) and meaning

- Construct: linguistic manifestation of grammatical expression licensed by a construction
Some Simple Examples

- Passive CxN licenses the following constructs:
  - The books were donated by Chuck to the library
  - The books were donated by Chuck.
  - The books were donated.

- Double_object CxN licenses the following constructs:
  - Ben sold Jerry the ice cream.
  - Paula gave Maria an apple.
Non-Canonical Grammatical Patterns

- idiomatic and irregular parts of language whose meanings are not predictable compositionally
- demonstrate frequency in text and centrality to language user’s knowledge
- exhibit structure, constraints, variation
Connection to Frame Semantics?

Frame Semantics and Construction Grammar are not just “sister” theories; rather they are sororal twins (Petruck 2014). Researchers in each theoretical framework must attend to the principles and goals of Frame Semantics and Construction Grammar.
“Beyond the Core” project (2008-2009)

- NSF-funded pilot project
- extend FrameNet to FrameNet Constructicon!
- collection of approximately 80 analyzed and annotated grammatical constructions
Road Map

- **Frame Semantics and FrameNet**
  - Frames, Frame Elements, Lexical Units
  - Annotation and Reports
  - Frame Relations and Frame Element Relations

- **Construction Grammar and Constructicons**
  - (Berkeley) Construction Grammar
  - The FrameNet Constructicon
  - Constructicon Development
Beyond the Core (2008-2009)

• Fillmore, Lee-Goldman, Rhodes (2012): “The kinds of constructions being collected and analyzed in the FrameNet Constructicon are mainly those that cannot be explained simply as instances of familiar constructions with ordinary lexical items.”

<table>
<thead>
<tr>
<th>Construction</th>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclamative CxN</td>
<td>What a beautiful boy!</td>
</tr>
<tr>
<td>“Clause-level” CxNs</td>
<td>These events in mind, he wrote a book.</td>
</tr>
<tr>
<td>Gapping CxN</td>
<td>He made no attempt to flirt with her nor she with him.</td>
</tr>
<tr>
<td>Adjective-as-noun CxN</td>
<td>She is a friend to the poor.</td>
</tr>
<tr>
<td>Verb-way CxN</td>
<td>He made her way to through the crowd.</td>
</tr>
<tr>
<td>Rate CxN</td>
<td>That plant grows four inches a day.</td>
</tr>
<tr>
<td>WXDY CxN</td>
<td>What’s that fly doing in my soup?</td>
</tr>
</tbody>
</table>
## Lexicon-Constructicon Analogues

<table>
<thead>
<tr>
<th>FrameNet Frames</th>
<th>FrameNet Constructicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>Construction</td>
</tr>
<tr>
<td>Instantiations of valence patterns</td>
<td>Construct</td>
</tr>
<tr>
<td>Frame Evoking Element (FEE) (Lexical Unit, LU)</td>
<td>Construction Evoking Element (CEE)</td>
</tr>
<tr>
<td>Frame Element (FE)</td>
<td>Construction Element</td>
</tr>
<tr>
<td>lexicographic annotation</td>
<td>constructicographic annotation</td>
</tr>
</tbody>
</table>
Construct: linguistic manifestation of grammatical expression licensed by a construction.

Construction: Constructions are the rules that license ‘new’ linguistic signs based on other linguistic signs (Fillmore 2012 et al.)

- Frames as semantic constructions
- LUs as lexical constructions
- Valence patterns as realization constructions
- Phrase types as building-block constructions
- Grammatical Functions as meta-realization constructions
- Control CxN and Supports (non-local argument realization)
Road Map

- **Frame Semantics and FrameNet**
  - Frames, Frame Elements, Lexical Units
  - Annotation and Reports
  - Frame Relations and Frame Element Relations

- **Construction Grammar and Constructicons**
  - (Berkeley) Construction Grammar
  - The FrameNet Constructicon
  - Constructicon Development
FrameNet Constructicon

- Proof of concept
- Add-on to FrameNet lexicon
- Inspiration for new constructicon development
  - Japanese
  - Brazilian Portuguese
  - Swedish
Constructicography
Inspired by ideas in Construction Grammar (Fillmore 1986 *inter alia*), annotated and automatically parsed relationships of cause and effect, in effect developing a construction of causals by including a variety of forms and discovering new ones in corpora.

Thanks!

Miriam R. L. Petruck
miriamp@icsi.berkeley.edu
The Multilingual FrameNet Project

Collin Baker (collinb@icsi.berkeley.edu)

International Computer Science Institute,
Berkeley, California

Multilingual FrameNet Tutorial 2018.08.20
at COLING, Santa Fe, NM
Overview

1. Paths to FrameNets
2. Coverage and Content
3. Alignment
4. Trial alignment Method and Results
Paths to FrameNets

Basic objectives and resources

- Practical vs. theoretical orientation
  - Berkeley FrameNet: Will Frame Semantics work in general? Will it help working lexicographers?
  - Spanish, German, Japanese FN, etc.: Will these frames work for ES, DE, JP? Are others needed?
  - Specific domains: Will FS work for sports texts, legal texts, environmental issues, dialog, argumentation, etc.? FN Brasil: Can we build an app for tourism?
Paths to FrameNets

Basic objectives and resources

- Practical vs. theoretical orientation
  - Berkeley FrameNet: Will Frame Semantics work in general? Will it help working lexicographers?
  - Spanish, German, Japanese FN, etc.: Will these frames work for ES, DE, JP? Are others needed?
  - Specific domains: Will FS work for sports texts, legal texts, environmental issues, dialog, argumentation, etc.? FN Brasil: Can we build an app for tourism?

- Lexicon vs. annotated corpus
  - Berkeley FN: build “Dictionary of the Future", annotate as documentation of usage; choose clear, simple examples
  - Later, began “full-text" annotation
  - Given a corpus, how well is it covered by FN? Annotate regardless of sentence complexity.
Paths to FrameNets

Existing Resources

- Existing corpora vs. DYI
  - BFN: British National Corpus, American National Corpus
  - SALSA: Tiger news corpus; no changes to parses
  - Japanese FN: Balanced Corpus of Contemporary Written Japanese (BCCWJ)
  - Spanish FN: built corpus from news, books, with a publisher, emphasizing New World Spanish

- Target language lexicons (national, free, commercial, bilingual)

- NLP tools: wide variation in target language tools: POS taggers, parsers, linked resources
Paths to FrameNets
Translation as part of the methodology

- **Korean FN:** (Hahm et al. 2018, Hahm et al. 2014) First hired translators to translate \( \sim 4,000 \) sentences from BFN, then fixed any errors. Later also translated from Japanese FN and projected annotation.

- **Hebrew FN:** (Hayoun & Elhadad 2016) Annotators chose English FN LUs in 200 frequent frames, found Hebrew translations (only LUs, not sentences).

- **Finnish FN:** (Lindén et al. 2017) The “annotated parts” of 80,000 BFN sentences were translated into Finnish in an early stage of development.
Coverage and Content

- The Team: ratio of Linguists to Computer Scientists
- General vs. domain-specific
- Frames and LUs vs. frames and word forms
- Spans vs. parse nodes vs. head words
- Annotate FEEs and FEs... and what else?
- Frame relations?
Coverage and Content
Coordination among FrameNets

- Distance from ICSI FN model
  - SALSA: "proto-frames"
  - French FN: merged frames (cf. Ruppenhofer et al. (2010))
  - FN Brasil: tourism frames

- Hard to update to new releases from ICSI just from the “diff” files
- How to feed back into Berkeley FN?
- Coordination between other FNs (including projection)
Coverage and Content

Availability of data produced

- Licensing and access methods
  - Licensing issues
  - Request to download/access website
  - Free to browse
  - Free to download
  - Available through API on the net
  - Part of linked linguistic open data on the cloud

- Types of data available
  - Frame names & descriptions
  - Lexical Units & definitions
  - Annotations without text
  - Annotated text
  - Frame & FE relations
  - Non-finalized or “problem” data
Linguistic Issues
Valences vs. constructions

- Define new frame or hypothesize construction?
- Parallel development of Construction Grammar
- Annotation of copulas and supports, presupposes constructions
- Negation frame and Conditional_occurrence.if.scon
- Sufficiency frame (*just enough* water to prevent them from sticking)
- Causatives in English and Japanese, among others ⇒ reorganization of frames?
Alignment across languages

Why align?

- Linguistic research
  - Comparative lexical semantics
  - Comparing valences/constructions across languages

- Implications for translation (human or machine)
  - Frames as an inter-lingua (Boas 2009)
  - Is translation "frame preserving"?
  - When and when not? Are frame shifts regular?
  - Can such knowledge help translation? (Čulo 2013)
Alignment across languages
Alignment by frames

Q: Since so many projects use Berkeley FN frames, why not just line up the frames?
A: Because of divergences from the Berkeley frames.
   - Is this the same frame? Same name? Translated name? Same ID?
   - Is this a "revised" frame? (changes in FEs, definitions)
   - Is this a new frame? (or in a later ICSI FN version?)
   - If new or revised, what is the relation to existing ICSI FN frames? Degree of relatedness? Type of relation?
Alignment across languages
Alignment by lexical units

- Frames can be aligned by measuring overlap of LUs across languages
- Sources for translation equivalents:
  - Bilingual dictionaries, i.e. Multilingual WordNet, BabelNet, UBY
  - Derive from parallel corpora, preferably sentence-aligned
  - Vector embeddings
- N.B. Partial solution to polysemy?—one sense per translation (cf. Carpuat 2009).
Distributional approaches

Introduction

- J. R. R. Firth (1957) and Zelig Harris (1954)
- Count words in windows near target word
- Different flavors: word2vec vs. GloVe
- Taking syntax into account: constituents or dependents
- Word forms vs. lemmas
Distributional approaches
Interpreting vectors

- Sparse binary vectors $d = 10,000+$
- Reduction to "standard" lengths $d = 40 \sim 300$
- Vector arithmetic, analogic reasoning
- Searching for meaning: QVEC (Tsvetkov et al. 2015) and SynEval (Köhn 2015, Köhn 2016)
- There are many ways to compose word vectors into vectors for phrases and sentences, including language models.
Trial Alignment
Multilingual distributional vector representations

- Our research on aligning is based on a shared, cross-linguistic vector space (cf. Hermann & Blunsom 2014 and Conneau et al. 2017)

- One approach would be to find the centroids of the frames in two languages and align those closest together in a shared space as discussed at this conference (Sikos & Padó 2018)

- Instead, we begin by formalizing the simple idea of translating LUs.

- Basically, we will say that a frame $X$ is *aligned with* a frame $Y$ when there are *enough* LUs associated with each that are good translations of each other.
Trial Alignment
Preparing the cross-lingual mappings

- For the first trial, we used frames and LUs from BFN 1.7 (the \textit{source}) and Spanish FN (the \textit{target}).
- We chose 200k word forms from the FN LUs for each language and retrieved vector representations for each of them from Facebook’s \texttt{fastText.cc} project (Bojanowski \textit{et al.} 2016).
- Note that this data includes vectors for \textbf{subparts} of words, between 3 and 6 characters long.
- Learned a linear map between them, yielding a joint-space representation, (cf. Conneau \textit{et al.} (2017).
- Removed source and target LUs for which no translation was found.
Trial Alignment

Alignment Algorithm

- For each word form $w_S$ of each LU in each source frame $f_S$, retrieve from the vector model the $N$ closest target translations, $w_T$
- For each of those translations, find the all the target frames $f_T^{(i)}$ in which they appear
- Each of those target word forms $w_T$ casts a "vote" for a mapping from $f_S$ to $f_T^{(i)}$
- If there are fewer than $k$ votes from a given LU to a given target frame, they are discarded. We have found by experimentation that with $N=10$ and $k=5$, we can avoid most incorrect mappings.
Example of translations (EN-ES)

- Building_subparts, (EN cloakroom, ⇒ ES comedor, habitación, cafetería, guardarropa, dormitorios, lavandería, sala, habitaciones, vestidor, dormitorio, aseos, sótano, palco, vestidores, antecámara, lavabos, boletería, entrepiso, retretes, vestíbulo)

- Building_subparts, (EN closet ⇒ ES armario, habitación, cama, sofá, clóset, sótano, apartamento, closet, ducha, bañera, armarios, desconcertado, lavabo, amordazado, vestidor, dormitorio, desconcertadamente, desván, ataúd, esconde)

- note subparts: <des consol ado>, <des consol ada mente>
Consoles
Consoles
Figure: Top 20 Berkeley FN frames with aligned Spanish FN frames
Figure: Top 80-150 Berkeley FN to Spanish FN frames
Discussion of Trial Alignment

A Few Comments

- Each of the previous slides ranks BFN frames by number of LUs.
- Many BFN LUs have no counterpart in Spanish FN (there are 13,631 LUs in BFN to 1,269 in SpFN and 1,073 BFN frames to 201 in the Spanish FN data).
- Moreover, we’re comparing BFN 1.7 to Spanish FN which is derived from BFN 1.5; frames have been changed, and some LUs moved.
- In the top 20 BFN frames seem to fan out more than the smaller frames, not surprisingly, given this algorithm.
In this preliminary work, we used only single-word LUs. Since multiwords constitute 12.8% of all LUs, we clearly need to find a way to handle them in the future.

We will learn alignments from other languages to BFN; as most of the other FrameNets have fewer frames and LUs, those should be cleaner alignments.

Surprisingly, even with a dictionary of 200k word forms, vectors were not found for a some of the FN word forms; we will try again with a larger dictionary.

We plan to align more languages by this method.
Recent developments and Future Research

- Crowdsourcing of frame discrimination (Chang et al. 2015, Hong & Baker 2011), (and FE annotation?)
- Learning about ambiguity by crowdsourcing (cf. Erk et al. 2013, Dumitrache et al. 2018)
- Working on a better display of valences (i.e. views on the FN database)
- Use Automatic Semantic Role Labeling: pre-annotation, post-annotation “in-filling”
- Frame relation induction? (e.g. Virk et al. 2016, Botschen et al. 2017)
Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1629989, which we gratefully acknowledge. This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.
Thank you for your attention!

https://framenet.icsi.berkeley.edu
Cross-lingual Semantics

Michael Ellsworth
(Adecco at Google)
mellswor@gmail.com
Outline

• Joint Multilingual FrameNet Annotation Task
  – Background and Software
  – First task: TED talk (English) and Translations
  – Cross-linguistic Framing Differences
  – Lessons

• New Work Necessitated by Task
  – Mental Spaces: Negation and Conditionals
  – Metaphor and Metonymy

• Frame-based Knowledge Representation?
Joint Multilingual Annotation: Background and Software

• Several frame semantic projects agreed to annotate in a shared task at ICCG9 in Brazil

• Initial conditions
  – a text with broad translation or comparable texts
  – no frames added beyond FN1.7
  – complete annotation of all frame-bearing words
  – words without frame in FN1.7 mapped to nearest

• Software
  – FrameNet Brasil's Web Tool
  – Hosted by Berkeley FrameNet
First Task:
Annotating TED Talk

• “Do Schools Kill Creativity”
  – most watched TED talk
  – very widely translated

• Text imported in multiple languages

• Different progress in annotation per project

• English and Portuguese most complete
Cross-linguistic Differences: English v. Portuguese

- Detailed comparison of only first 30 sentences
- Spurious differences:
  - differing definitions of “complete” annotation
  - differing mappings for words not in FN1.7
  - annotation mistakes
  - translation “errors”
- True differences:
  - less than 10%
  - most obvious for spatial language
Lessons

- Need clearer definition of complete annotation
  - are coordinating conjunctions frame-bearing?
- Need process for addressing other spurious differences
  - mistakes discovered by comparison to parallel anno
  - recognize multiple valid/equivalent nearest frames
- Need more participants! This is fun!
New Work

• Future joint tasks will include creating new frames

• A broad definition of complete annotation includes very frequent phenomena with only a partial treatment in FrameNet
  - Negation
  - Conditionals
  - Metaphor
  - Metonymy
Frame-based Knowledge Representation?

- Given results so far, frames very similar in translations
- Opens possibility of representing texts more language neutrally via frames
- Desirable properties of frames as KR
  - Frames capture language, language captures what people think is important
  - Better alignment between human and machine representations and models
Automatic frame-semantic role labeling

Swabha Swayamdipta
Carnegie Mellon University
Outline

1. **Task of frame-SRL**

2. General Methodologies

3. Taking Stock

4. Looking forward: Multilingual Extensions
Frame-semantic Role Labeling (frame-SRL)

1. Target Identification
   - Hoover Dam
   - played (play.v)
   - a major (major.a)
   - role
   - in preventing (prevent.v)
   - Las Vegas
   - from drying up (dry up.v)
   - Preventing_cause
   - PERFORMERS_AND_ROLES
   - IMPORTANTANCE

2. Frame Identification
   - PERFORMER
   - ROLE
   - PERFORMANCE
   - IMPORTANCE
   - FACTOR
   - THWARTING
   - PROTAGONIST
   - ENTITY
   - ACTION
   - BECOMING_DRY
A close relative: PropBank SRL

1. Target Predication Identification
2. Frame Sense Identification
3. Frame-Elements Argument Identification

Example from Johansson & Nugues (ACL 2008)

Chrysler plans new investment in Latin America

plan.01 investment.01
A little bit of history

• Pioneered by Gildea & Jurafsky (CL 2002) on an earlier version of FrameNet.

• Development of PropBank (Kingsbury & Palmer, 2002; Palmer et. al., 2005)

• CoNLL shared tasks expedited the development on PropBank-style SRL.

Frame-SRL as Graph Induction

• Treating various components as graphs

  • Nodes: Tokens / Spans in the sentence

  • Node Labels: Targets

  • Edges: Between Targets and Frame-Elements

  • Edge Labels: Roles of frame-elements
Target Identification

• Predict “semantically salient” tokens as targets in the sentence.

• Challenge: No simple POS tag based bijection, unlike in PropBank, where most verbs are targets.

• Multi-word expressions also considered valid targets. About 4% of the corpus.

• FN 1.5: On an average, about 6 targets per sentence.
Frame Identification

• Given a target (lexical unit) token in the sentence, identify the frame evoked by it.

• On an average, about 2 frames per target - lexical unit.

• Lexical units play a critical role here, because of the mapping between lexical units and frames.

  ▶ Errors in identifying lexical units / targets directly affect frame identification.
Argument Identification

• Given a target and the frame it evokes, identify all the spans in the sentence which are arguments to the frame, and their respective labels (frame-elements)

• Primary difference between PropBank SRL and Frame SRL argument identification task:
  • PropBank argument labels are uniform across predicates, even though divergent definitions of each label is permitted
1. Task of frame-SRL

2. How to build a frame-SRL model?

3. Taking Stock

4. Looking forward: Multilingual Extensions
Frame-SRL task

- Most common approach: Supervised Learning
- Unsupervised approaches (Titov & Klementiev, 2012)
- Semi-supervised approaches (Das et. al., CL 2014)
- Data (78 full-text annotated documents in FN 1.5)
  - Train (47) / Dev (8) / Test (23)
- Learning with exemplar annotations (Kshirsagar et. al., NAACL 2015)
Model architectures

• Linear models - most models prior to 2015
  • May use distributional representations

• Non-linear / neural models
  • Deep bidirectional, highway (Zhou & Xu, ACL 2015; He et. al. ACL 2017)
  • Transformers (Tan et. al., AAAI 2018; Strubell et. al., 2018)
Target ID models

- Heuristic methods (Johansson & Nugues, IWSE 2007; Das et. al., CL 2014)

- Neural methods (Swayamdipta et. al., 2017)

- Challenges:
  - Data sparsity
Target Identification Heuristics

- **have** was retained only if had an object,
- **be** was retained only if it was preceded by **there**, 
- **will** was removed in its modal sense,  
- **of course** and **in particular** were removed, 
- the prepositions **above, against, at, below, beside, by, in, on, over, and under** were removed unless their head was marked as locative, 
- **after** and **before** were removed unless their head was marked as temporal, 
- **into, to, and through** were removed unless their head was marked as direction, 
- **as, for, so, and with** were always removed,  
- because the only sense of the word of was the frame PARTITITIVE, it was removed unless it was preceded by **only, member, one, most, many, some, few, part, majority, minority, proportion, half, third, quarter, all, or none, or it was followed by all, group, them, or us,**  
- all targets marked as support verbs for some other target were removed.

**Johansson & Nugues (IWSE, 2007)**
**Das et. al. (CL 2014)**
Frame ID models

- With features from syntax (Das et. al., CL 2014)

- With distributional semantics (Hermann et. al., ACL 2014)

- Neural approaches (Swayamdipta et. al., 2017; Yang et. al., EMNLP 2017; Peng et. al., NAACL 2018)

- Out of domain frame identification (Hartmann et. al., EACL 2017)
Arg ID models

- Span classification
- Sequence labeling
- Each frame treated independently
Role of syntax

- Traditional feature design
- Continuous valued features
- Constraints during decoding
- Heuristics for potential argument identification
Learning strategies

- Log-linear objectives
- Cost-augmented objectives
- KL divergence between objectives of different kinds
Multitask learning

- With syntax (Swayamdipta et. al. 2017; Swayamdipta et. al. EMNLP 2018)
- With PropBank semantics (Kshirsagar et. al. NAACL 2015; Fitzgerald et. al. EMNLP 2015)
- With semantic dependencies (Peng et. al. NAACL 2018)
Inference

- Constrained inference (Das et. al., CL 2014):
  - Non-repetition of core frame-elements
  - Non-overlap of argument spans
  - Pairwise inclusions / exclusions

- Greedy inference (Henderson et. al., CL 2013; Swayamdipta et. al., 2017)
  - Typically, only respects non-overlapping arguments

- Joint inference of frames and frame-elements (Yang et. al. EMNLP 2017; Peng et. al. NAACL 2018)

- Joint inference of predicates, senses and arguments (Labeled Span Graphs, He et. al., ACL 2018)
1. Task of frame-SRL

2. General Methodologies

3. Taking Stock

4. Looking forward: Multilingual Extensions
ARK Syntactic & Semantic Parsing Demo

Here you can test two open-source linguistic structure analysis tools from the Noah’s ARK research group at Carnegie Mellon University:

- **TurboParser** performs tokenization, part-of-speech tagging, and syntactic dependency parsing. (more...)
- Using the syntactic parse as input, **SEMAFOR** produces a FrameNet-style analysis of semantic predicate-argument structures. (more...)

**Details**

```
I need some good caffeine to wake me up!
```

**Syntactic Dependency Parse from TurboParser [0.8s]**

Tree View (explanation)

**CoNLL Format**

**Frame-Semantic Parse from SEMAFOR [0.1s]**

Grid View (explanation)

- Horizontal  ○  Vertical
Applications

- Question Answering (Shen & Lapata, EMNLP 2007)
- Information Extraction (Surdeanu et. al. ACL 2003)
- Text-to-scene generation (Coyne et. al., COLING 2012)
- Dialog systems (Chen et. al., ASRU-IEEE 2013)
- Social network extraction (Agrawal et. al. EACL 2014)
- Knowledge Extraction from Twitter (Søgaard et. al., AAAI 2015)
- Machine Translation (Marchegianni et. al., ACL 2018)
Biggest challenges

• Coverage
  • FrameNet+ (Pavlick et. al. ACL 2015)
  • Augmentation via Paraphrases (Rastogi & Van Durme, Workshops at ACL 2015)

• Domain Adaptation
  • Distributional semantics (Hartmann et. al. EACL 2017)
Outline

1. Task of frame-SRL
2. General Methodologies
3. Taking Stock
4. Looking forward: Multilingual Extensions
Multilingual FrameNet

- Potential applications
  - Translation using semantics as pivot
  - Transfer learning for downstream applications such as knowledge | information | relation extraction in low-resource languages
Extensions to other languages

- Unsupervised induction (Titov & Klementiev, ACL 2012)

- Any language frame-semantic parsing (Johannsen et. al. 2015)
  - 9 languages in 2 domains
  - Using word-word translation
  - Inter-annotator agreement issues stemming from automatic target identification through word-word translation
Polyglot models

- Training data from pairs of languages merged (Mulcaire et. al. ACL 2018)

- Challenge: Differences in annotation schemes across languages.

- Multilingual word embeddings, learned from cross-lingual alignments (Ammar et. al., 2016)

- Maximum benefit reported for low-resource languages such as Catalan, when combined with English.
Challenges again...

- Coverage
- Domains
- Alignments
Summary

- Part I: Task
  - a. Target Identification
  - b. Frame Identification
  - c. Frame-Element Identification

- Part II: General Methodologies
  - a. Model Architectures
  - b. Role of Syntax
  - c. Learning Strategies: Pipeline vs Joint

- Part III: Taking Stock

- Part IV: Multilingual Extensions