Corpus-based analysis and annotation of constructions

Russell Lee-Goldman and Russell Rhodes

{rleegold, russell_rhodes}@berkeley.edu

University of California, Berkeley

Frames and Constructions
31 July–2 August 2009
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Purpose of talk

- Report on a year-long project creating a database of English constructions or an English *constructicon*
  - Parallel to FrameNet in terms of goals and approach
  - Building a resource for people and, potentially, machines
  - *Not* building a full-fledged Construction Grammar
- Provide a representative sample of the current coverage of the *constructicon*
- Briefly discuss potential applications for such a resource

The project was conceived of and initiated by Chuck Fillmore. There would be no results to report if not for him.
FrameNet (FN; http://framenet.icsi.berkeley.edu/) is an on-line lexical resource for English, based on frame semantics and supported by corpus evidence.

- **Frames**
  - “[A] script-like conceptual structure that describes a particular type of situation, object, or event along with its participants and props” (Ruppenhofer, et al.)
- **Frame Elements**
  - Participant (or prop) roles of the frames are identified and defined
  - Words are grouped based on the frame they evoke
  - A *Lexical Unit* (LU) is the pairing of a word and frame
- **Annotation**
  - Sentences which represent clearly the meaning and combinatory possibilities of Lexical Units are collected from a corpus and the frame elements are annotated
A simple illustration

Self-motion

- The Self_mover, a living being, moves under its own power in a directed fashion.
- Frame Elements: Self_mover, Source, Path, Goal, ..., Speed
- \([Self\_mover \text{Most cars}] \text{CRAWL} [\text{Path along}] [\text{Speed at 15 miles an hour}]\).

FrameNet Annotation

<table>
<thead>
<tr>
<th>Layer</th>
<th>Most</th>
<th>cars</th>
<th>crawl</th>
<th>along</th>
<th>at</th>
<th>f i f t e e n</th>
<th>miles</th>
<th>s a n</th>
<th>hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GF</td>
<td>Ex t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>N P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RLG and RR (UC Berkeley) Construction annotation Fillmore Fest 6 / 24
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Limitations of the lexicon

Lexical annotation alone does not fully capture the meaning of sentences.

Often constructions are needed for full understanding.

- **Morphological**
  - “Where the sad-eyed prophet says that no man comes...”

- **Phrasal**
  - *When weddings were weddings*, and the scullery maid would faint if she were seen by her ladyship in the wrong part of the house not attending to her duties, there were rules of etiquette and runic codes to cover every contingency.
  - For years he had been **close friends** with Kevin McCarthy and his wife.
Example revisited

Example
Most cars crawl along at fifteen miles an hour

- Entire expression describes a speed—but how?
- Speed is a complex notion (distance, time), and this information is in the sentence
- Annotating from the perspective of any single word won’t reveal that complexity
- Special syntax

We need a RATE construction (or maybe even a SPEED construction) which describes the constituent parts and the whole they produce, and its semantic and pragmatic import.
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Goals of a constructicon

Goals:
- Repository of constructions
- Human and machine readable
  - Useful to both linguists and NLP researchers
- *Not* a full-fledged construction grammar, though it should be useful for building one

Parts:
- List of constructions
- Construction definitions
  - Description of syntactic, semantic, and pragmatic aspects of constructions and their parts
- Annotation of sentences illustrating the constructions ("constructicographic annotation")
We identify and annotate the following:

- **CONSTRUCT**
  - A phrase licensed by the rules of a construction
- Phrase type of the construct
- **CONSTRUCT ELEMENTS** (CE; components of the construct) and their phrase types
- A special **CONSTRUCTION-EVOKING ELEMENT** (CEE)
  - Not all constructions have such an element
  - Analogous to the target LU in FN annotation
- Relevant features of the context
**Rate.speed**

- The Rate.speed construction indicates speed of an Entity in terms of Distance that would be traveled for each period of (iterated) Time.
- Construct Elements: Entity, Distance, Time
- \[ Entity \text{ Most cars} \] crawl along at \[ Distance 15 \text{ miles} \] \[ Time \text{ an hour} \].

**Constructional Annotation**

<table>
<thead>
<tr>
<th>LU/Oxn</th>
<th>Layer</th>
<th>Most cars</th>
<th>crawl along at</th>
<th>Distance</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate.speed CE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate.speed GF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate.speed PT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate.speed CstrPT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate.speed CEE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate.speed GovX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RLG and RR (UC Berkeley)
Construction annotation
Fillmore Fest
A richer example

Let\_alone

<table>
<thead>
<tr>
<th>Construction</th>
<th>None of these arguments is notably strong, let alone conclusive.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEE</td>
<td>None of these arguments is [CEE let alone] conclusive.</td>
</tr>
<tr>
<td>CEs</td>
<td>None of these arguments is [First_conjunct notably strong], let alone [Second_conjunct conclusive].</td>
</tr>
<tr>
<td>Construct</td>
<td>None of these arguments is [Let_alone notably strong, let alone conclusive_AJP].</td>
</tr>
<tr>
<td>Contextual Element</td>
<td>[Trigger None] of these arguments is notably strong, let alone conclusive.</td>
</tr>
</tbody>
</table>
### Parallels between constructions and frames (more practical)

<table>
<thead>
<tr>
<th>Frame</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame element</td>
<td>Construct Element</td>
</tr>
<tr>
<td>Target</td>
<td>Construct</td>
</tr>
</tbody>
</table>

### Parallels between lexical items and constructions (more theoretical)

<table>
<thead>
<tr>
<th>LU</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>evokes a frame</td>
<td>(may) evoke a frame</td>
</tr>
<tr>
<td>licenses FEs</td>
<td>(may) license FEs</td>
</tr>
<tr>
<td>has several valences</td>
<td>has several valences</td>
</tr>
</tbody>
</table>
Our work

In our one year project, we created:

- Descriptions of 73 constructions and their CEs
- Annotation of 50 of these constructions
- About 1500 sentences

Adjective_as_nominal.definite
Adjective_as_nominal.people
As.role
Attributive_degree_modification
Be_recip
Comparison_equality
Comparison_inequality
Degree_so
Deictic_dayname_inverted
Dimension_conjunction
Exocentric_adjectival_compound
Gapping_Integrated_appositive
Inversion_with_preposed_element
Let_alone
Location_in_calendar_subunit
Location_in_calendar_unit
Measurement_plus_adjective
Measurement_plus_prepositional_phrase
Ones_very_eyes
Open_interrogative.non-subject
Own_right
Proper_name_embellishments
Rate.cost_time
Rate.frequency
Rate.mileage
Rate.speed
Rather_than_coordination
Shared_Completion
Stripping
SAI.closed_interrogative
SAI.conditional
SAI.emphatic_negative_imperative
SAI.exclamation
SAI.optative
SAI
Subject-predicate
Superlative
Supplement_ascriptional
Supplement_specificational
Tagged_sentence.canonical
Tagged_sentence.subjectless
Tautology.nostalgia
Uniqueness
Way_manner
Way_means
Way_neutral
What-with_absolutive
With_absolutive
Types of constructions

Some constructions do not evoke frames:

- **Gapping:** ...the singular includes the plural and the plural the singular...
  
  \[
  \text{[Before the singular]} \ [Gapped\ includes] \ [After the plural] \ [\text{Conj and}] \ [After the singular]\]

- **Shared_completion:** He is clearly familiar with and fond of that cat.
  He is clearly \[\text{Sharer familiar with}\]
  \[\text{Connector and}\]
  \[\text{Sharer fond of}\] \ [Completion that cat].
Types of constructions

Other constructions do evoke frames and appear with FEs:

- Make one’s way evokes the Motion frame:
  \[\text{Theme } \text{We} \] \[\text{Manner}_\text{verb} \text{sang} \] \[\text{CEE } \text{our way} \] \[\text{Path } \text{across Europe} \]

- Be friends with evokes the Reciprocity frame:
  For years \[\text{Party}_1 \text{he} \] had been \[\text{Modifier}_\text{close} \] \[\text{Head}_\text{noun} \text{friends} \] \[\text{Party}_2 \text{with Kevin McCarthy and his wife} \].

- Rate constructions, the comparative and superlative constructions, ...
Types of constructions

Some constructions include fixed lexical material

- *one’s own right*: $[\text{in\_own\_right\_entity}\text{Linguistics}]$, like psychology, has grown up, and flown the nest of $[\text{Established\_entity}\text{philosophy}]$, to become $[\text{Property}\text{a science}] [\text{CEE in its own right}]$.

- *one’s way* in make one’s way, -er or *more* in comparative, -est or *most* in the superlative, *let alone*, ...

This fixed material is either the CEE or makes up part of the CEE
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Potential Applications

- Researchers
  - A resource for linguists who are concerned with providing “full-coverage” of English
- Education
  - Aid for second language teachers and students
- Computational resources
  - Improving natural language processing (NLP) applications
    - Semantic role labeling
    - Parsing
Outline

1. Introduction
2. Limitations of FrameNet
3. Constructicon
4. Potential Applications
5. Conclusions
Conclusions

- Reported on the creation of an English constructicon
- Discussed similarities and differences between the constructicon and a frame-based lexicon
- Provided a sample of the current contents of the constructicon
- Briefly discussed potential applications for such a resource
Acknowledgments

The very existence of the project is due to Chuck Fillmore and a grant from the National Science Foundation.

Thanks to Collin Baker, Chris Oei, Hiroaki Sato, Michael Ellsworth, and everyone else at FrameNet for technical and analytic assistance.

And thanks to the volunteers and organizers of FillmoreFest, especially Eve Sweetser and Miriam Petruck.