

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

Introduction

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

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* Most cars] CRAWL [*Path* along] [*Speed* at 15 miles an hour].

FrameNet Annotation

Layer	M	o	s	t	c	a	r	s	c	r	a	w	l	a	l	o	n	g	a	t	f	i	f	t	e	e	n	m	i	l	e	s	a	n	h	o	u	r					
FE	S	e	l	f	_	m	o	v	e						P	a	t	h	S	p	e	e	d																				
GF	E	x	t												D	e	p	D	e	p																							
PT	N	P													P	P	P	P																									
Other																																											
Verb																																											
Sent																																											

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 Construction**
- 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”)

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

A richer example

Let_alone

CEE	None of these arguments is notably strong, [<i>CEE</i> let alone] conclusive.
CEs	None of these arguments is [<i>First_conjunct</i> notably strong], let alone [<i>Second_conjunct</i> conclusive].
Construct	None of these arguments is [<i>Let_alone</i> notably strong, let alone conclusive _{AJP}].
Contextual Element	[<i>Trigger</i> None] of these arguments is notably strong, let alone conclusive.

Adapting FN to Constructicon

Parallels between constructions and frames (more practical)

Frame	Construction
Frame element	Construct Element
Target	Construct

Parallels between lexical items and constructions (more theoretical)

LU	Construction
evokes a frame	(may) evoke a frame
licenses FEs	(may) license FEs
has several valences	has several valences

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_absolute
 With_absolute

Types of constructions

Other constructions do evoke frames and appear with FEs:

- Make one's way evokes the Motion frame:
 [*Theme* We] [*Manner_verb* sang] [*CEE* our way] [*Path* across Europe]
- Be friends with evokes the Reciprocity frame:
 For years [*Party_1* he] had been [*Modifier* close] [*Head_noun* friends]
 [*Party_2* 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*: [*In_own_right_entity*Linguistics], like psychology, has grown up, and flown the nest of [*Established_entity*philosophy] , to become [*Property*a science] [*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.