Computational semantics
– Introduction –

Annette Hautli

Universität Konstanz

Machine Language Processing

Motivation

Semantics of natural language

"You know, somebody actually complimented me on my driving today. They left a little note on the windscreen; it said, 'Parking Fine'. So that was nice."

Tim Vine (English actor)

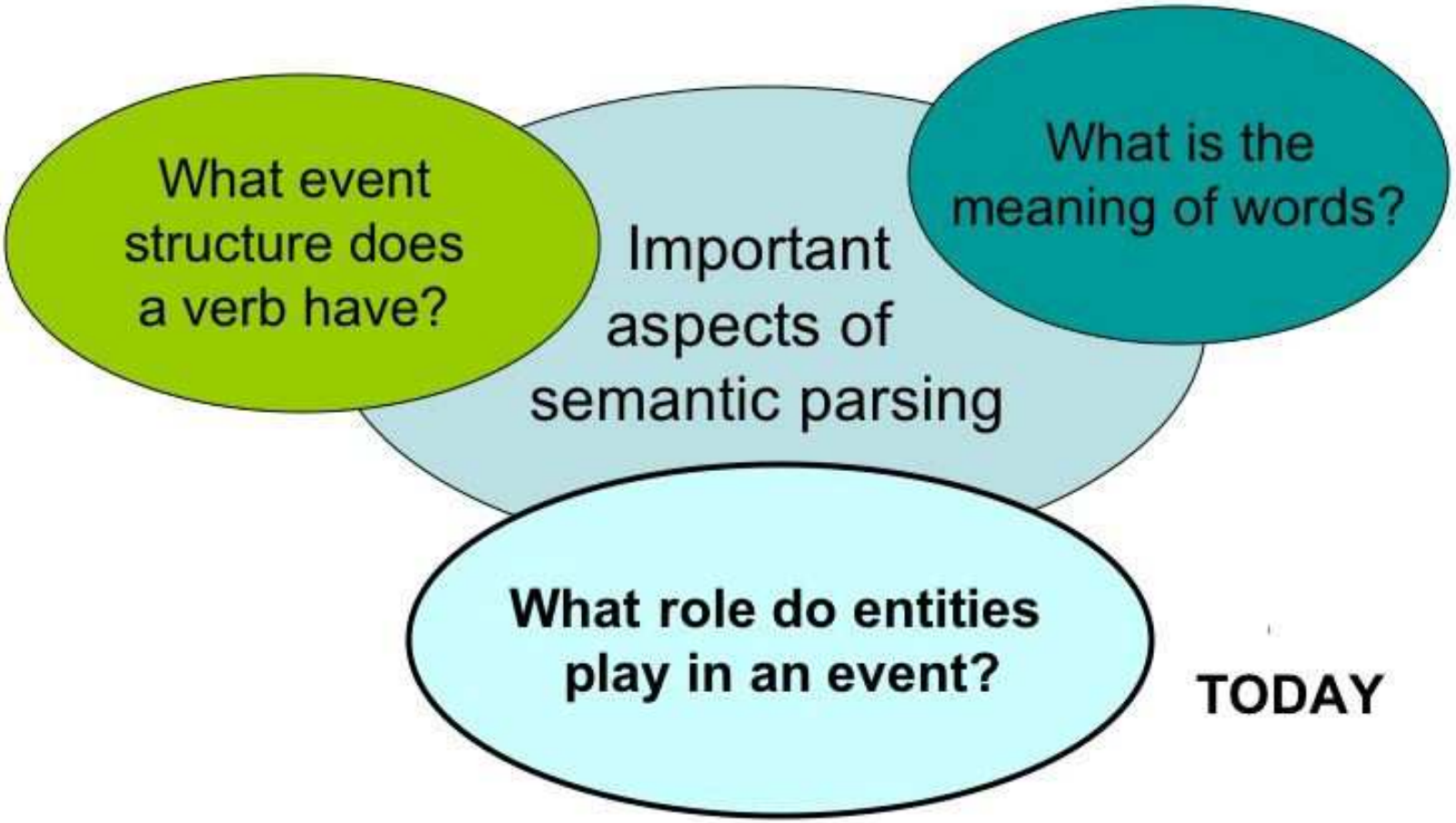
Motivation

- Syntactic parsers have become more and more robust
- Automatic semantic analysis has come into the reach of NLP applications

Central question of semantic NLP:

Who did What to Whom,
and How, When and Where?

Overview


What event structure does a verb have?
What is the meaning of words?
Important aspects of semantic parsing
What role do entities play in an event?
TODAY

Outline

Syntax

Argument structure

- Each verb has a unique meaning associated with the event it denotes
- This meaning is intrinsically linked to the number of participants in the event

→ Verbs are classified as intransitive (one participant), transitive (two participants) and ditransitive (three participants)

Syntax

Intransitives

(1) a. Ram slept under a tree.

b. The girl sang all afternoon.

Transitives

(2) a. John broke the vase.

b. The farmer sowed the rice.

Ditransitives

(3) a. John sent Mary books.

b. The man put the money on the table.

Syntax

Stochastic syntactic analysis (parser trained on PennTreebank)

"I enjoy teaching in Konstanz."

Syntax

Stochastic syntactic analysis (parser trained on PennTreebank)

"I enjoy teaching in Konstanz."

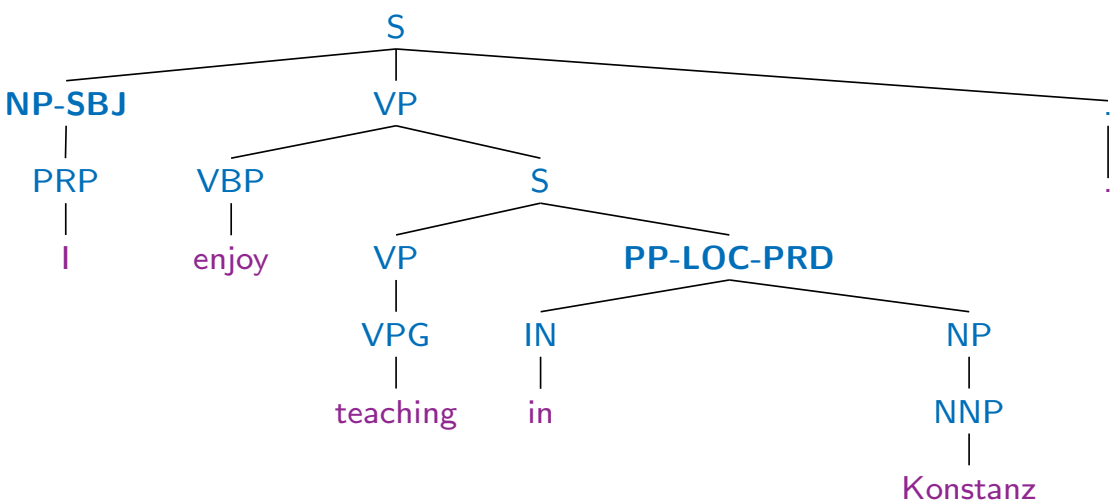
(S (NP-SBJ (PRP I)) (VP (VBP enjoy) (S (VP (VPG teaching)
(PP-LOC-PRD (IN in) (NP (NNP Konstanz))))))(. .))

Syntax

Stochastic syntactic analysis (parser trained on PennTreebank)

"I enjoy teaching in Konstanz."

(S (NP-SBJ (PRP I)) (VP (VBP enjoy) (S (VP (VPG teaching) (PP-LOC-PRD (IN in) (NP (NNP Konstanz))))))(. .))

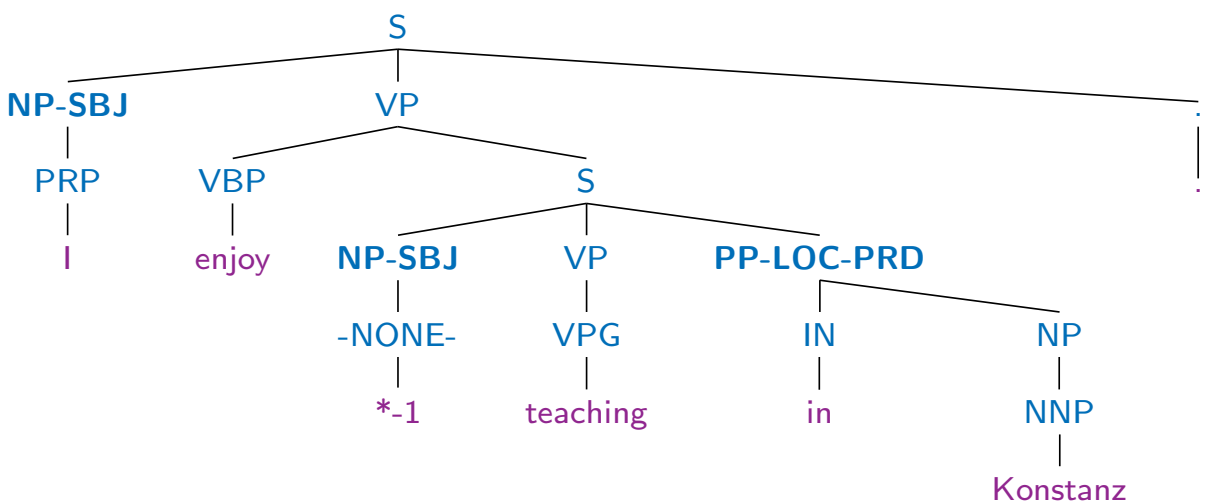


Syntax

Stochastic syntactic analysis (parser trained on PennTreebank)

"I enjoy ___ teaching in Konstanz." — long-distance dependency

(S (NP-SBJ (PRP I)) (VP (VBP enjoy) (S **(NP-SBJ (-NONE- *-1))** (VP (VPG teaching) (PP-LOC-PRD (IN in) (NP (NNP Konstanz)))))) (. .))



Syntax

Rule-based deep grammars (English XLE grammar)

"I enjoy teaching in Konstanz."

```
[PRED    'enjoy<[-5-SUBJ:I], [-5-XCOMP:teach]>'
 SUBJ    [PRED  'I'
          NTYPE [NSYN pronoun]
          CASE nom, HUMAN +, NUM sg, PERS 1, PRON-TYPE pers]
 XCOMP   [PRED    'teach<[-5-SUBJ:I]>'
          SUBJ    [-5-SUBJ:I]
          ADJUNCT { [PRED  'in<[-4-OBJ:Konstanz]>'
                     OBJ   [PRED  'Konstanz'
                            CHECK [_LEX-SOURCE morphology]
                            NTYPE [NSEM [PROPER [LOCATION-TYPE city, PROPER-TYPE location]]
                                   NSYN proper]
                            CASE obl, NUM sg, PERS 3]
                     PSEM  {loc}
                    -4 PTYPE sem] }
          CHECK   [_SUBCAT-FRAME V-SUBJ]
          TNS-ASP [PERF -_, PROG +_]
          PASSIVE -, VTYPE main]
 CHECK   [_SUBCAT-FRAME V-SUBJ-XCOMPprog]
 TNS-ASP [MOOD indicative, PERF -_, PROG -_, TENSE pres]
-5 CLAUSE-TYPE decl, PASSIVE -, VTYPE main]
```

NLP and meaning

Some questions in semantic NLP:

1. In what way do syntactic arguments participate in an event?
   - assignment of thematic roles to syntactic arguments $\rightarrow$ linking theory
2. Is there a way of automatically assigning these roles?

NLP and meaning

1. Thematic roles

Claim: thematic roles are consistent across different syntactic realizations

	John	broke	the vase.
Syntax:	Subject	Verb	Object
Thematic roles:	Agent		Patient

NLP and meaning

1. Thematic roles

Claim: thematic roles are consistent across different syntactic realizations

	John	broke	the vase.
Syntax:	Subject	Verb	Object
Thematic roles:	Agent		Patient
	The vase	broke.	
Syntax:	Subject	Verb	
Thematic roles:	Patient		

NLP and meaning

1. Thematic roles

Claim: thematic roles are consistent across different syntactic realizations

	The vase	was broken.
Syntax:	Subject	Verb
Thematic roles:	Patient	

NLP and meaning

1. Thematic roles

Claim: thematic roles are consistent across different syntactic realizations

	The vase	was broken.	
Syntax:	Subject	Verb	
Thematic roles:	Patient		
	The vase	was broken	by John
Syntax:	Subject	Verb	PP
Thematic roles:	Patient		Agent

Outline

Basics

- Assumption: Sentence consists of a verb and one or more noun phrases
- Each noun phrase is related to the verb with a particular *case* relationship (*case* = thematic role)
- Each case relationship occurs only once in a sentence
- Verbs are classified according to the available case relations in the sentence

Examples

(4) a. [John]$_{SUBJ}$ broke the window.
Agent Theme

b. [The hammer]$_{SUBJ}$ broke the window.
Instrument Theme

c. [John]$_{SUBJ}$ broke the window with a hammer.
Agent Theme Instrument

Examples

(5) a. [John]$_{SUBJ}$ broke the window.
Agent Theme

b. [The hammer]$_{SUBJ}$ broke the window.
Instrument Theme

c. [John]$_{SUBJ}$ broke the window with a hammer.
Agent Theme Instrument

d. #[John and the hammer]$_{SUBJ}$ broke the window.
Agent Instrument Theme

e. #[The hammer]$_{SUBJ}$ broke the window with a chisel
Instrument Theme Instrument

Fillmore's case notions

Agent (A)

The case of the perceived instigator of the action, typically animate.

John opened the door.

Instrumental (I)

The case of the inanimate force or object causally involved in the action or state.

The key opened the door.

Dative (D)

The case of the animate being affected by the state or action.

He gave *his brother* the ball.

Fillmore's case notions

Factitive (F)

The case of the object or being resulting from the state or action of the verb.

The mother baked *a cake*.

Locative (L)

The case of the location or spatial orientation of the state or action of the verb.

I flew to *Kathmandu*.

Objective

The case of things which are affected by the state or action of the verb.

He gave his brother *the ball*.

Fillmore's case relations

- Each verb has a specific array of cases, for example:
  - *to run*: [Agent ___]
  - *to remove/to open*: [Agent ___ Objective]
  - *to give*: [Agent ___ Dative Objective]
- BUT: verbs can appear with more than one case frame

Fillmore's case relations

Task #1

What are the case frames for the constructions below?

(6) a. The wind opened the door.

b. She gave the idea a second thought.

c. The water runs.

d. The thought of Ben killed her.

Role	**Example**
Agent	*John* opened the door.
Instrumental	*The key* opened the door.
Dative	He gave *his brother* the ball.
Factitive	The mother baked *a cake*.
Locative	I flew to *Kathmandu*.
Objective	He gave his brother *the ball*.

Fillmore's case relations

Task #2

What are the problems of Fillmore's system (particularly in NLP applications)?

Outline

Basics

- Two core roles: PROTO-AGENT and PROTO-PATIENT
- Role types are not discrete categories but rather cluster concepts
- "Different degrees of membership" in a role type
- Arguments entail features of the proto-roles

Proto-roles

PROTO-AGENT properties
a. volitional involvement in the event or state e.g. *John is polite to Mary.*
b. sentience (and/or perception) e.g. *John sees/fears Mary.*
c. causing an event or change of state in another participant e.g. *John's loneliness causes his unhappiness.*
d. movement (relative to the position of another participant) e.g. *Water filled the boat.*
(e. exists independently of the event named by the verb) e.g. *John needs a new car.*

Proto-roles

PROTO-PATIENT properties

a. undergoes change of state or location
   e.g. *John moved the rock.*
b. incremental theme
   e.g. *John filled the glass with water.*
c. causally effected by another participant
   e.g. *Smoking causes cancer.*
d. stationary relative to movement of another participant
   e.g. *The bullet entered the target.*

(e. does not exist independently of the event, or not at all)
   e.g. *John built a house.*

Proto-roles

- Less prototypical Agent: Experiencer
  - Sentience, causation, movement
  - *Sally felt the heat.*
- Even less prototypical Agent: Instrument
  - Causation, movement
  - *The knife cut through the flesh.*
- Less prototypical Patient: Theme
  - Incremental theme, change
  - *Ben watched the game.*

Links between grammatical relations and thematic roles

Argument selection principle

- Argument with the highest number of Proto-Agent properties is the subject
- Argument with the highest number of Proto-Patient is the direct object

Links between grammatical relations and thematic roles

Task #3

What role properties do the arguments in the following constructions correspond to?

(7) a. Captain Nemo sank the ship with a torpedo.

b. A torpedo sank the ship.

c. The ship sank.

(8) a. John fears thunder.

b. Thunder scares John.

Role hierarchy

- The argument selection principle determines a thematic role hierarchy
- " $\rangle$ " means "outranks for subject"

$$\text{Agent} > \left\{ \begin{array}{l} \text{Instrument} \\ \text{Experiencer} \end{array} \right\} > \text{Patient} > \left\{ \begin{array}{l} \text{Source} \\ \text{Goal} \end{array} \right\}$$

- Strong agents outrank strong patients for subjecthood
- Instruments and experiencers outrank patient-like argument

Outline

Definitions adapted from http://www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsASemanticRole.htm

- **Agent**: a person or thing who is the doer of an event
- **Patient/Theme**: affected entity in the event; undergoes the action
- **Experiencer**: receives, accepts, experiences, or undergoes the effect of an action
- **Stimulus**: the thing that is felt or perceived
- **Goal**: place to which something moves, or thing toward which an action is directed.
- **Recipient** (sometimes grouped with Goal)

- **Source** (sometimes grouped with Goal): place or entity of origin
- **Instrument**: an inanimate thing that an Agent uses to implement an event
- **Location**: identifies the location or spatial orientation of a state or action
- **Manner**: how the action, experience, or process of an event is carried out.
- **Measure**: notes the quantification of an event

Let's annotate some data!

Agent
Patient
Experiencer
Stimulus
Goal
Recipient
Source
Instrument
Location
Manner
Measure

1. Mary hid in the cupboard.
2. Sam broke the vase.
3. The vase broke.
4. Carla fears snakes.
5. She reached Mumbai early.
6. The mother gave the child some candy.
7. Her dress cost 10 €.
8. Cookies make great desserts.

Problems with thematic roles

- No fixed, finite set of thematic roles definable
- The delimitation of roles is problematic
- Conceptual domains can cut across role distinctions

One way to deal with this in NLP:

- adopt a small set of roles (PropBank)

Outline

Facts

- Section of Penn Treebank 2 annotated with thematic roles
- 112917 annotated examples
- 3257 unique verbs
- Core arguments are numbered, optional arguments receive a label
- Primary goal of developing an annotated corpus as training data for supervised machine learning systems
- Facilitates experiments of the sort that dominate NLP currently

Frame file

Frame: increase.01

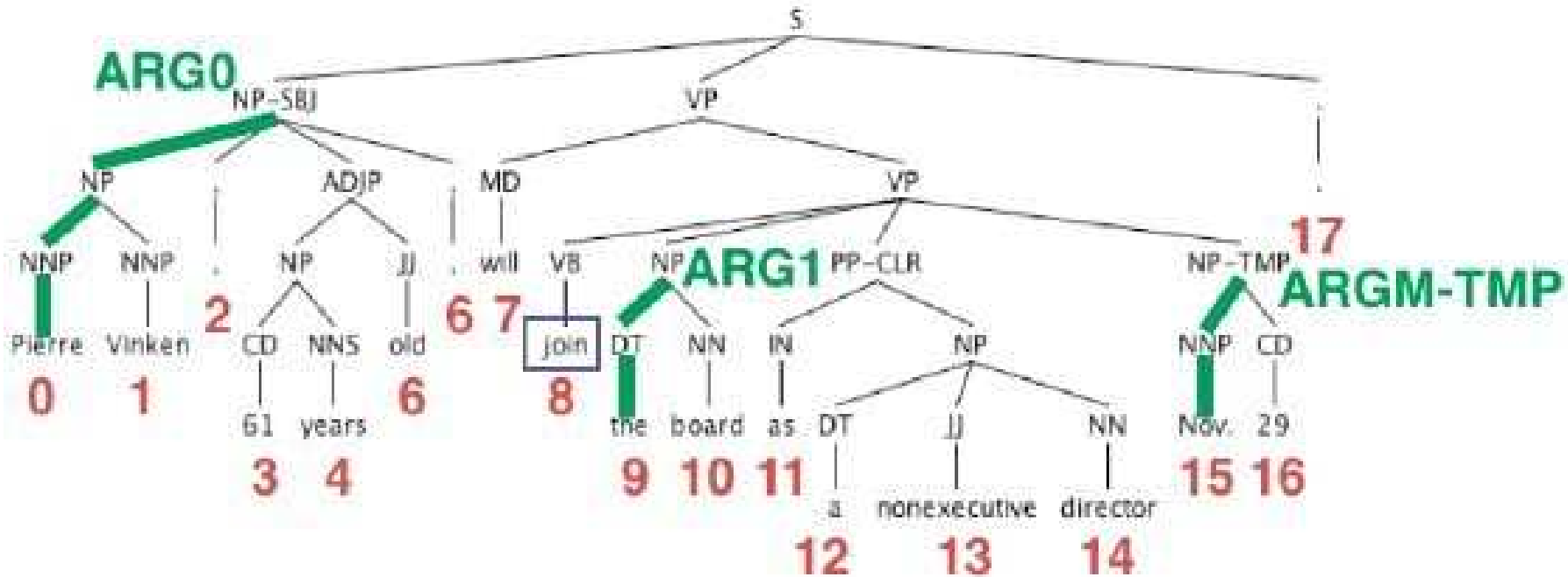
- name: go up incrementally
- vncls: 45.4 45.6
- ARG0 causer of increase (vntheta: Agent)
- ARG1 thing increasing (vntheta: Patient)
- ARG2 amount increased by, EXT or MNR (vntheta: Extent)
- ARG3 start point (vntheta: –)
- ARG4 end point (vntheta: –)

Examples

1. [ARG0 The Polish government] [rel increased] [ARG1 home electricity charges] [ARG2-EXT by 150%].
2. [ARG1 The nation's exports] [rel increased] [2-EXT 4%] [4-2 to $50.45 billion].
3. [ARG1 Output] will be [2-MNR gradually] [rel increased] .

taken from Christopher Potts on Semantic Role Labeling

Frame file



taken from Christopher Potts on Semantic Role Labeling

PropBank argument labels

Label	
rel	the verb
ARGA	causative agent
ARGM	adjuncts
ARG0	generally subj
ARG1	generally dobj
ARG2	generally iobj
.	
.	
.	

Label	
EXT	extent
DIR	direction
LOC	location
TMP	temporal
REC	reciprocal
PRD	predication
NEG	negation
MOD	modal
ADV	adverbial
MNR	manner
CAU	cause
PNC	purpose not cause
DIS	discourse

taken from Christopher Potts on Semantic Role Labeling

Propbank

Virtues

- Full gold-standard parses
- Different levels of annotation for one corpus (syntax, thematic roles)

Drawbacks

- labels ARG2-5 are overloaded (VerbNet and FrameNet provide more fine grained role labels)
- WSJ is too domain specific and too financial
- What about argument-taking nouns and adjectives?

Outline

Outline

Overview

Semantic role labeling (SRL)

The task of automatically finding the semantic roles for each predicate in the sentence.

Overview

Semantic role labeling (SRL)

The task of automatically finding the semantic roles for each predicate in the sentence.

The girl$_{Agent}$ ate the cake$_{Theme}$.

Overview

Semantic role labeling (SRL)

The task of automatically finding the semantic roles for each predicate in the sentence.

The girl$_{Agent}$ ate the cake$_{Theme}$.

- View from Information Extraction
  - basic event structure
  - "Who did what to whom, how when and where?"
- View from Computational Linguistics
  - identify predicate-argument structure
  - identify the arguments of a sentence, assign semantic labels that describe their roles

SRL slides taken from Jurafsky and Martin (chapter 20)

Overview

- Boost in interest in SRL in the last decade
  - Availability of large enough annotated corpora (PropBank, FrameNet)
- Typically, SRL is done from parses
  - Syntax-semantics interface
  - Need of parsed data or parser
- Complex systems, technical difficulties, theoretical linguistic challenge

Outline

Typical approach

Step 1: Identification: which phrases are role-bearing?

- Necessary for real-world tasks
- Rather syntactic task
- Constituents that can be an argument
- Role-bearing phrases need not necessarily be constituents
  $\rightarrow$enormous search space
- Arguments are more easy to identify than adjuncts (70-90% against 60%)

Typical approach

Step 2: Classification: what role do role-bearing phrases play?

- More semantic task
- Highly dependent on the set of roles
- Importance of selectional restrictions (lexical semantics)
- Enormous search space

Step 3: Evaluation: how accurate is the labeling?

- Very tricky to get right
- Are some argument types more important than others?
- Are some mis-classifications worse than others?

Gildea and Jurafsky (2002)

Semantic role labeling system

- Seminal work on assigning semantic roles to a corpus
- Features needed for their labeling system:
  - Lexical (predicate, subcategorization frame, head word of the constituent)
  - Syntactic (phrase type of constituent, POS of headword, path)
  - Other (voice, linear position)

How good can we get?

- last 5 years: 9 evaluations in several competitions
  - CoNLL, SemEval, SensEval, several years
- SRL on PropBank
  - given predicates, identify+label arguments
  - $\approx$ 80% F-score on Wall Street Journal
  - $\approx$ 70% Brown corpus
  - results on unseen predicates, WSJ: $\approx$ 70% (low generalization)
- realistic setting (SemEval 2007): overall $\approx$ 45% F-score
  - identify predicates and arguments + label them in running text; FrameNet roles

Problems

Current systems

- Complex (integration of syntax and semantics, identification, ranking, combination of classifiers...)
- Slow!
- Syntax: need parser $\rightarrow$ degrades performance $\rightarrow$ bottleneck
- Domain dependency: consistent degradation of 10% of f-score when training on WSJ and testing on Brown

Wrap-up

Who did What to Whom,
and How, When and Where?

1. In what way do syntactic arguments participate in an event?
   - Several theoretical approaches to thematic roles are available
   - Concept is hard to formalize
2. Is there a way of automatically assigning these roles?
   - Yes, shown as early as Gildea and Jurafsky (2002)
   - Performance and speed still need to be improved

Wrap-up

How do you call Santa's little helpers?

Wrap-up

How do you call Santa's little helpers?

Subordinate Clauses.

Happy Christmas!

from walkinthewords.blogspot.com