

Semantic Role Labeling with Cornerstone and Jubilee

A decorative graphic consisting of several horizontal lines of varying lengths and colors (teal, light blue, white) extending from the right side of the slide.

Outline

- Overview of the tools
- Framefile creation with Cornerstone
 - Numbered Args versus ArgMs
 - Multiword expressions
- Annotation with Jubilee
 - Numbered arguments
 - Choosing Arg0 or Arg1
 - Modifiers
 - Light verb annotation
 - Noun annotation

Overview: Cornerstone and Jubilee

- Developed for PropBank annotation
- Supply simple, general-purpose labeling of semantic roles
- Used to provide consistent argument labels across different syntactic realizations
- Assign general functional tags to all modifiers or adjuncts to the predicate

Different verb senses . . .

- Have different subcategorization frames
- PropBank assigns coarse-grained senses to verbs
- PropBank “framesets,” lexical resource
- New senses, or “rolesets” are added only when the syntax and semantics of a usage are distinct
- Annotators use “frame files” to assign appropriate numbered arg structure

Framefile lists framesets

- *Mary left the room*
- *Mary left her daughter-in-law her pearls in her will*

Frameset **leave.01** "move away from":

Arg0: entity leaving

Arg1: place left

Frameset **leave.02** "give":

Arg0: giver

Arg1: thing given

Arg2: beneficiary

Overview: Cornerstone and Jubilee

- Rolesets designed using Cornerstone, which creates .xml files
- Cornerstone .xml files feed into Jubilee
- Jubilee also uses phrase structure annotation as a base (Penn Treebank, Marcus et al., 1993)
- Annotation of text done with Jubilee

General argument assignments

- ARG0 agent
- ARG1 patient
- ARG2 instrument, benefactive, attribute
- ARG3 starting point, benefactive, attribute
- ARG4 ending point
- ARGM modifier

Frame for *break*

Frameset **break.01** “break, cause to not be whole”:

Arg0: breaker

Arg1: thing broken

Arg2: instrument

Arg3: pieces

Function tags for all verb modifiers or adjuncts (not predicate-specific)

- TMP - when? *yesterday, 5pm on Saturday, recently*
- LOC - where? *in the living room, on the newspaper*
- DIR - where to/from? *down, from Antartica*
- MNR - how? *quickly, with much enthusiasm*
- PRP/CAU -why? *because ... , so that ...*
- REC - himself, themselves, each other
- GOL - end point of motion, transfer verbs? *To the floor, to Judy*
- ADV - hodge-podge, miscellaneous, “nothing-fits!”
- PRD - this argument refers to or modifies another: *...ate the meat raw*

Break.01 revisited

- No need for numbered arguments for components not specific to the predicate

ArgM-TMP

- **On the first day of his new job**, he broke the light bulb over his desk.

ArgM-CAU

- He broke the light bulb **out of spite**.

ArgM-MNR

- He **accidentally** broke the light bulb.

Arg3

- He broke the light bulb **into a thousand pieces**.

Numbered Arg or ArgM?

- The distinction is closely tied to the argument-adjunct distinction
- Fairly generous in assigning argument status to something

Arg0 Arg1 Arg2 Arg3

- [I] bought [the car] [from Pat] [for \$50.]

Arg0 Arg1 ArgM-PRP

- [He] betrayed [his country] [for \$50.]

Assigning Arg numbers

- Generally, most central or most necessary arguments are assigned lower numbers
- The more peripheral, the higher the arg number
- Arg0 reserved for Dowty-like proto-Agent
- Arg1 not specifically reserved for proto-Patient

Open Cornerstone

- `java -jar cornerstone-1.35.jar en susanb`
- Go to File – Open
- Browse to Jubilee – resource – frameset
- Click on a file ending with `.xml`

Let's make a framefile: Scribble

- I scribbled her a note.
- They're adorable until they scribble on your property.
- He scribbles nonsensical phrases.
- He always scribbles on the covers of his textbooks.
- The code was scribbled on the back of an envelope.
- They scribbled away for hours.
- As they scribble and paint, they move in different ways.

Scribble

- I scribbled her a note.
- They're adorable until they scribble on your property.
- He scribbles nonsensical phrases.
- He always scribbles on the covers of his textbooks.
- The code was scribbled on the back of an envelope.
- They scribbled away for hours.
- As they scribble and paint, they move in different ways.
- **Why did they scribble out the face?**

PB framefile for *scribble*

Scribble.01: Draw/write untidily

- **Arg0:** *scribbler*
- **Arg1:** *thing scribbled*
- **Arg2:** *benefactive*
- I scribbled him a note saying I didn't know any of the Trivial Pursuit answers.

Twist

- He was watching me as I twisted Evan's drumsticks in my hands.
- She twisted her ankle.
- What if she twisted the truth to fit in?
- You don't twist the top off; you just pull.
- Follow the road as it twists left and up the hill.
- This rope is long, and it twists better.
- The movie "Colorblind" will twist your mind.

PB twist framefile

Twist.01: to change form, shape, etc.

- **Arg0:** *cause of change, agent*
- **Arg1:** *entity twisted*
- The husky blond guide with the Aggie cap twists his face in mock fury .
- You twist people's trust.

Similar verb

- **run.04** , *range*, *extend*
- **Roles:**
- **Arg1:** *logical subject, thing ranging*
Arg2: *EXT, distance run*
Arg3: *start point*
Arg4: *end point*
Argm: *medium*

You try the verb *bank*

- John banks at First Trust Bank.
- Why don't we close down the poker game and bank the \$16 billion?
- I would only bank my paycheck in an institution guaranteed by the FDIC.
- They banked the full amount of \$4,000.

- She banked the plane sharply to avoid the helicopter.
- The plane banked gently, following the river.
- As it got closer, it banked to the right.

- They **banked up** the road with soil.
- The new owners **banked up** the brook.

Official PropBank framefile

- **bank.01** , *put (money) in the bank*
- *Arg0: putter, agent*
Arg1: money
Arg2: bank
- **bank.02** , *turn like an airplane*
- *Arg0: pilot*
Arg1: aeroplane
- **Bank.03**, *create an embankment*
- *Arg0: creator*
Arg1: embankment or location
Arg2: substance of embankment

Multiword expressions

- Verb-particle constructions
 - English: a roleset within a framefile
 - Chinese: separate framefile (new lemma)
- Idiomatic expressions
 - No expression-specific rolesets or framefiles
 - One generic Idiomatic Expression roleset used by annotators
- Light verb usages
 - Complicated 2-stage annotation procedure

Open Jubilee

- version # user ID
- Java -jar jubilee-2.16.jar -u browns