Summer 2008 Training Refresher

Verb Sense Annotation Project
Different Types of Data

- Wall Street Journal
- Brown Corpus
- Broadcast News
- Broadcast Conversation
Wall Street Journal

☑ Watch out for financial jargon
  ☑ If you don’t understand what a passage is talking about, ASK!
  ☑ You don’t have to understand perfectly to figure out POS and the general sense
  ☑ There are sentence fragments in the WSJ - you can still tag it as a verb if that’s the correct POS
Brown Corpus: Literature

✓ Watch out for metaphorical extensions of different senses - you can still tag these as the same sense
✓ There are sentence fragments in Brown - you can still tag it as a verb if that’s the correct POS
✓ Brown is the most frequent place to find new senses not in the grouping. Be sure to email verb.annotation@gmail.com if you choose None of the Above and suggest a sense
Disfluencies: <disfluency>

Repetition: a verb is repeated several times before the full sentence is given

- Reread the last repetition of the verb, with all of its arguments present, first. You may have to grab one argument from the beginning of the whole string and then the other argument(s) from the end.

- Tag each repetition of the verb, in order, IDENTICALLY.

- If you have any questions, please email verb.annotation@gmail.com.
Broadcast News and Conversation Data Guidelines

- Truncation: a sentence is ended prematurely or a new sentence begins in the middle of an existing utterance
  - Tag the verb with a sense only if you are sure, either via present arguments or context, that you can tell the meaning of the verb
  - If you don’t know what the verb means because there’s too little information, tag it X, NOT None of the Above
Linguistic Jargon: Constituents

- Constituent is a fancy word for Part of Speech (POS). It lets us talk about phrases.

- Constituents we’re concerned about:
  - Noun Phrase (NP)
  - Verb Phrase (VP)
  - Adjectives, Adjective Phrases (AP)
Constituents and Grammar Rules

S -> NP VP
VP -> V NP
NP: I
V: love
NP: working
(PP -> P NP)
P: at
NP: the office
NP V NP (PP{P NP})
I love working (at the office.)

Part of speech is defined by POSITION; that’s why ‘working’ in this sentence is a noun, but in “She was working yesterday” it’s a verb.
Constituents as Arguments

- NP V NP: I love NASCAR.
  - Verb has 2 arguments; transitive
- NP V NP NP (ditransitive): I gave him the ball.
  - Verb has 3 arguments
- NP V NP PP{P NP}: I put the salad in the refrigerator.
  - Verb has 2 arguments and an oblique; transitive

When we talk about finding arguments, we mean a verb’s subject and object(s).
- NP V PP{P NP}: I went to the store
  - Verb has one argument; this is intransitive
Linguistic Jargon: Syntax

✓ Transitivity: transitive/intransitive/ditransitive
  ✓ NP1 V; NP1 V NP2; NP1 V NP3 NP2
  ✓ I run; I run the dog; I run Mary the dog (I run the
dog to Mary.)

✓ NP numbering is linear
✓ Constituent labels appear in linear order
✓ Template: Optional prepositional phrase
  ✓ (PP{P NP[feature]})
Linguistic Jargon: Semantics

✓ Verbs and nouns often carry semantic roles/features in their syntactic description

✓ Major semantic roles:
  ✓ Agent: I ate the apple.
  ✓ Patient: She hit the bear.
  ✓ Theme: We discussed politics.
  ✓ Goal (Target): The teacher threw a lot at us.
  ✓ Location: I took the trash out to the alley.
  ✓ Instrument: I cut the apple with a knife.
  ✓ Recipient/Beneficiary: I gave the book to my mother.
  ✓ Topic: It was a big step for the company.
    ✓ Old information in a stream of discourse.
  ✓ Focus: The really great thing was the new factory, though.
    ✓ New information in a stream of discourse.
Linguistic Jargon: Semantics

- Semantic Features: used either on their own or to modify semantic roles
  - State: The band stirred the crowd into a frenzy: NP1[agent] STIR NP2[patient] PP{P into NP[state]}
  - Concrete: Able to be directly perceived by the senses: She felt the horse.
  - Abstract: Not able to be directly perceived by the senses: She felt the sadness of that place.
  - Durative: Takes up an amount of time: She left for an hour.
  - Punctual: Occurs instantaneously: She left at noon.
Other Helpful Terms

- **Auxiliary**: a verb that is not the main verb in the sense; carries tense/aspect/mode but no semantic information.
  - I have taken courses at CU.

- **Verb Particle Construction (VPC)**: verb plus a particle (preposition without a complement), demonstrated by movement:
  - She put off her work/She put her work off.

- **Realis**: talking about reality: It rains when Sally goes to the store.

- **Irrealis**: talking about possibility: It would rain were Sally to go to the store.

- **Nominalization**: making a verb into a noun, using -ing suffix or a determiner (article)
  - RUN: The run through town is fun/Running is fun.

- **(Past)-Participle**: -ed or irregulars; may be an adjective
- **Finite**: verb is not participle or infinitive; carries tense information
- **Non-finite**: verb is participle or infinitive; no tense information
Predicate Adjectives: NP BE ADJ

✓ Sometimes, predicate adjectives and verbs (especially in the passive) look the same:
  ✓ The house was finished.
  ✓ The house was finished by the workers.
  ✓ The house was finished by Friday.

✓ Not just the passive, but many non-locative PP phrases (with the exception of “of”) seem to give a verbal reading. What is the distribution?
  ✓ The house was built at noon. [temporal] - ambiguous
  ✓ X The house was built of stone. [attribute] - ADJ
  ✓ X The house was built on stone. [locative/attribute] - ADJ
  ✓ ? The house was built on the hill (yesterday). - with temporal modifier - VERB. Without modifier - ADJ.

✓ What are the diagnostics we could use to tell the difference between adjectives and verbs in the past participle?

✓ How can we be consistent in our annotation?
Adjective/Verb Diagnostics

1. Context: can we get an [agent] argument from the surrounding text?
   ✓ The President gave the order and the army was based in Nepal.
   ✓ The tornado was based in Wichita.

2. Can you make it part of a relative clause?
   ✓ The company, [that was] based in Texas, moved to Tennessee.
     ✓ If it is a reduced relative clause, then the past participle is not a verb

3. The more we build up the post-’verbal’ phrase/clause, the more it reads like a verb, so be careful.
   ✓ The man is trapped. - ADJ
   ✓ The man is trapped in the cellar with an axe. ADJ
   ✓ ? The man was trapped in the cellar yesterday with an axe. VERB
   ✓ ? The house was built before the war. ADJ

4. Unless part of a conventionalized phrase (e.g. born yesterday), past participles with more “punctual” temporal modifiers (although these are all durative) should be tagged as verbs.