

NLP 202: Topics in Syntax

Jeffrey Flanigan

Winter 2023

University of California Santa Cruz

jmflanig@ucsc.edu

Plan For Today

Linguistics

- Syntactic roles
- Active vs passive sentences
- Semantic roles
- Levin verb classes

NLP

- Lexicalized PCFGs

If time: Dependency syntax

- Pat throws a crayon at us.
- They throw a crayon at us.
- *Pat throw a crayon at us.
- *They throws a crayon at us.

“Pat” and “they” are special: they are the **subject**. In English, a present-tense verb agrees in number with the subject.

Syntactic roles: Subject vs Object

- Pat threw the ball.

Pat is the subject.

Ball is the object.

Syntactic roles: Subject vs Object

Quechua (spoken in the Andes mountains):

- Juan Pedro-ta maqan. “Pedro hits Juan.”
Juan Pedro-SUBJ maqan.
- Pedro-ta Juan maqan. “Pedro hits Juan.”
Pedro-SUBJ Juan maqan.

The morpheme “-ta” marks the subject. Marking the syntactic role of nouns using morphemes is called **case marking**.

Syntactic roles: Subject vs Object

In English, the subject comes before the verb and the object comes after.

→ Syntactic roles are specified using **word order**.

In Quechua, the subject is marked with a special morpheme, and the object is unmarked.

→ Syntactic roles are specified using **case marking**.

In general, syntactic roles are specified using word order or case marking.

Active vs passive sentences

- Pat threw the ball.
- You can make the object the subject:
The ball was thrown.

Active vs passive sentences

- Pat threw the ball. **active sentence**
- You can make the object the subject:
The ball was thrown. **passive sentence**
“was” is an auxiliary verb. In English, “was” sometimes denotes a passive sentence, sometimes denotes past tense, or both.

Active vs passive sentences

- Pat threw the ball. **active sentence**
- You can make the object the subject:
The ball was thrown. **passive sentence**
“was” is an auxiliary verb. In English, “was” sometimes denotes a passive sentence, sometimes denotes past tense, or both.
- The original subject can be included as an prepositional phrase (called an **oblique**):
The ball was thrown by Pat.

Active vs passive sentences

Japanese:

- Taro-ga Hanako-o nagut-ta
Taro-SUBJ Hanako-OBJ hit-PAST “Taro hit Hanako.”
- Hanako-ga nagut-**rare**-ta
Hanako-SUBJ hit-PASS-PAST “Hanako was hit.”

Active vs passive sentences

Japanese:

- Taro-ga Hanako-o nagut-ta
Taro-SUBJ Hanako-OBJ hit-PAST “Taro hit Hanako.”
- Hanako-ga nagut-**rare**-ta
Hanako-SUBJ hit-PASS-PAST “Hanako was hit.”
- Hanako-ga Taro-ni nagut-**rare**-ta
Hanako-SUBJ Taro-OBL hit-PASS-PAST “Hanako was hit by Taro.” **Can include the original subject as an oblique.**

Transitive vs intransitive verbs

Some verbs allow an object, and some do not.

- I eat.
- I eat pizza.
- I sleep.
- *I sleep the bed.
- I sleep on the bed. (Expressed as a prepositional phrase)

Verbs that take objects are called **transitive verbs**. Verbs that do not allow objects are called **intransitive verbs**.

Ditransitive verbs

Some verbs allow a second object (**ditransitive** verbs).

- I gave a book to Mary.
- SUBJ gave OBJ OBL
- I gave Mary a book.
- SUBJ verb OBJ OBJ2

second NP with no preposition → called OBJ2

Passive:

- A book was given to Mary by me.
- SUBJ PASS verb OBL OBL

Original subject and object can be expressed with obliques.

Dummy Subjects

- It is large. (“It” is a pronoun that refers to something.)
- It is windy. (“It” is a dummy subject, because English requires a subject for sentences that aren’t commands.)
- It tends to get dark early around here.

- I bought a car from my friend for \$5,000.
- **BUYER THING-BOUGHT SELLER PRICE**

More general semantic roles (sometimes called **thematic roles**):

- Sherlock heard a piercing scream.
- **EXPERIENCER STIMULUS**
- Pat hit the ball with a bat.
- **AGENT PATIENT INSTRUMENT**
- John opened the lock with a key.
- **AGENT PATIENT INSTRUMENT**

INVENTORY OF SEMANTIC ROLES:

AGENT: causer or initiator of events

EXPERIENCER: animate entity which perceives a stimulus or registers a particular mental or emotional process or state

RECIPIENT: animate entity which receives or acquires something

BENEFICIARY: entity (usually animate) for whose benefit an action is performed

INSTRUMENT: inanimate entity used by an agent to perform some action

THEME: entity which undergoes a change of location or possession, or whose location is being specified

PATIENT: entity which is acted upon, affected, or created; or of which a state or change of state is predicated

STIMULUS: object of perception, cognition, or emotion; entity which is seen, heard, known, remembered, loved, hated, etc.

LOCATION: spatial reference point of the event. The **LOCATION** role includes the sub-types **SOURCE**, **GOAL**, and **PATH**, which respectively describe the origin (or beginning-point), destination (or end-point), and pathway of a motion

ACCOMPANIMENT (or COMITATIVE): entity which accompanies or is associated with the performance of an action

- George opened the door. subject = person (with volition) causing the event
- George opened the door. subject = person causing the event

Verb alternations

Some verbs allow certain variations, aka **alternations**

Consider “break” vs “hit.”

Instrument alternation:

- The boy broke the window (with a ball).
- The boy hit the window (with a ball).

Body-part alternation:

- I broke his leg. / *I broke him on the leg.
- I hit his leg. / I hit him on the leg.

Another alternation:

- Janet broke the vase / *Janet broke at the vase.
- Carla hit the door. / Carla hit at the door.

- bend, **break**, crack, fold, shatter, split, snap... allow the same alternations as break.
- bash, bump, **hit**, kick, pound, slap, strike, tap, whack... allow the same alternations as hit.

The BREAK-type verbs are different semantically than HIT-type verbs:

- BREAK-type verbs: Involve a change of state
- HIT-type verbs: Surface contact, without requiring a change of state.

For example:

- # The rocks broke the windshield, but luckily it wasn't damaged. # means semantically incoherent
- The rocks hit the windshield, but luckily it wasn't damaged.

Levin Verb Classes

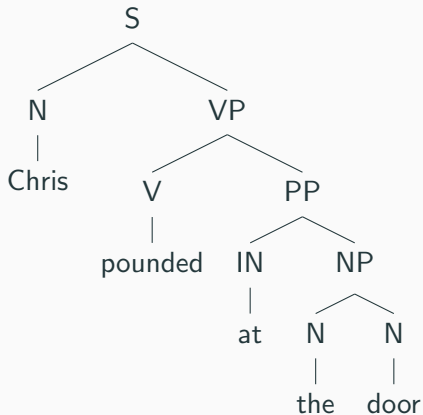
- You can categorize verbs by the alternations they allow, and look for semantic similarities to explain the phenomena
- Beth Levin did this, and published a book of her findings (“English Verb Classes and Alternations” (Levin, 1993))
- The classes she found are called **Levin verb classes**

Take-aways

- Different verbs allow different prepositional phrases, alternations, etc
- Syntax (which alternations a verb allows) is interrelated with semantics

This is why word vectors work so well for parsing (more on this later)

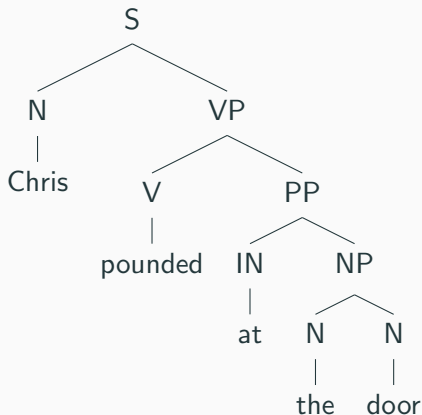
What assumptions does a CFG/PCFG make?



Let's try to fix this problem.

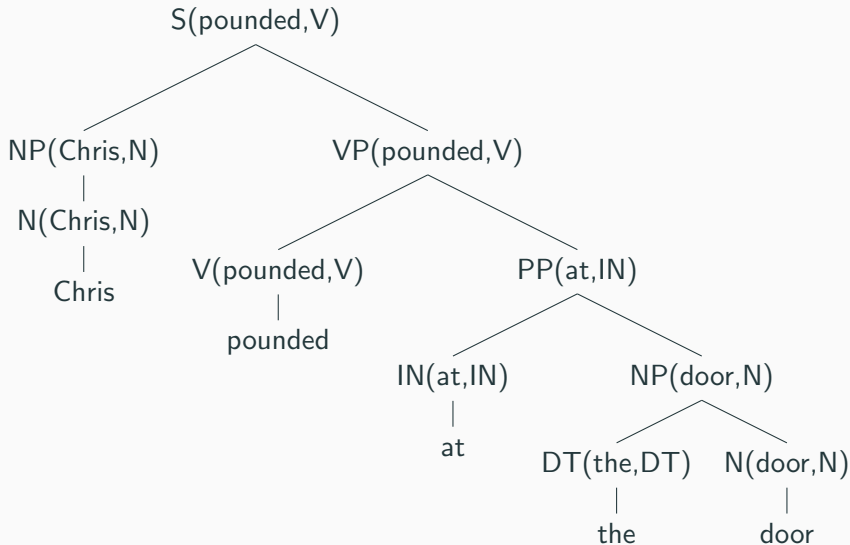
Problems with PCFGs

- The words in a PCFG have little influence on the overall parse structure.



Lexicalized PCFGs

Annotate each node with its **head** + POS tag of the head



The **head** of a constituent is the main word for the phrase.

- John sees the [red truck]. **Head of “red truck” is truck**
- The [five quiet people] see the red truck. **Head of “five quiet people” is people**

Lexicalized PCFGs

- Annotate each node with its **head** + POS tag of head
- We can't have a rule for each fine-grained production — e.g. $P(S^{(\text{saw}, \text{VBD})} \rightarrow NP^{(\text{man}, \text{NN})} VP^{(\text{saw}, \text{VBD})})$
- Different models make different **independent assumptions** to make this quantity tractable (Collins 1999, Charniak 1997)