Syntax and morphosyntax

May 20, 2020

Low Resource NLP Bootcamp

Lori Levin

Outline

- Categories of words (Parts of Speech)
  - Discreteness and universality of parts of speech
- Morphosyntax and Interlinear Glossed Text (IGT)
  - Collaborative Exercise: The North Wind and the Sun
- Syntax
  - Collaborative Exercise: Universal Dependency Treebank

Part 1: Parts of Speech

Words fall into categories (e.g., noun, verb, adjective, preposition, postposition, adverb, determiner) that help us describe grammar.

Four English Noun Phrases and two ways to describe them

The book
A book
This student
These assignments

One way to describe the four Noun Phrases:

- An English noun phrase can start with the words *the, a, this, that, these,* or *those*.
- The second word can be *book, student, cat, teacher, bus,* or *assignments.*
- *These* goes with books, students, teachers.
- *A* goes with book, student, teacher.
- *This* goes with book, student, teacher.
- *The* goes with book, books, student, students, teacher, teachers, etc.

A different way to describe the four noun phrases:

- *Determiner* and *noun* are types of words.
- Determiners can be singular or plural.
  - Singular determiners: a, the, this, that
  - Plural determiners: these, those, the
- Nouns can be singular or plural
  - Singular nouns: book, student, assignment, teacher, cat, bus
  - Make a noun plural by adding –s/-es to the singular noun
- You can make an English noun phrase out of a determiner followed by a noun.
- The determiner and the noun must have the same number (singular or plural).

Four English Noun Phrases and two ways to describe them

The book
A book
This student
These assignments

One way to describe the four Noun Phrases:

- An English noun phrase can start with the words *the, a, this, that, these,* or *those*.
- The second word can be *book, student, cat, teacher, bus,* or *assignments.*
- *These* goes with books, students, teachers.
- *A* goes with book, student, teacher.
- *This* goes with book, student, teacher.
- *The* goes with book, books, student, students, teacher, teachers, etc.

A different way to describe the four noun phrases:

Lexicon

- *Determiner* and *noun* are types of words.
- Determiners can be singular or plural.
  - Singular determiners: a, the, this, that
  - Plural determiners: these, those, the
- Nouns can be singular or plural
  - Singular nouns: book, student, assignment, teacher, cat, bus
  - Make a noun plural by adding –s/-es to the singular noun

Grammar

- You can make an English noun phrase out of a determiner followed by a noun.
- The determiner and the noun must have the same number (singular or plural).

Four English Noun Phrases and two ways to describe them

The book
A book
This student
These assignments

Linguists use parts of speech to describe grammar more efficiently, e.g., by separating the lexicon and the grammar.

Linguists use parts of speech to show similarities and differences between languages.

A different way to describe the four noun phrases:

Lexicon

- *Determiner* and *noun* are types of words.
- Determiners can be singular or plural.
  - Singular determiners: a, the, this, that
  - Plural determiners: these, those, the
- Nouns can be singular or plural
  - Singular nouns: book, student, assignment, teacher, cat, bus
  - Make a noun plural by adding –s/-es to the singular noun

Grammar

- You can make an English noun phrase out of a determiner followed by a noun.
- The determiner and the noun must have the same number (singular or plural).

A traditional approach to parts of speech

- Verbs denote actions
- Nouns denote entities
- Adjectives denote states
- Adverbs denote manner
- Prepositions denote location
- Determiners specify

My son's second grade grammar homework

- I **wished** for a pony.
- I have an **idea**.
- Sam **is** smart.

- Why is "wish" an action but "idea" is a thing?
- Why is "is" an action?

A traditional approach to parts of speech

- Verbs denote actions
- Nouns denote entities
- Adjectives denote states
- Adverbs denote manner
- Prepositions denote location
- Determiners specify

My son's second grade grammar homework

- I **wished** for a pony.
- I have an **idea**.
- Sam **is** smart.

- Why is "wish" an action but "idea" is a thing?
- Why is "is" an action?

The problem is that these definitions are not operationalized so that different people can always come to the same conclusion about whether something is a verb, noun, adjective, etc.

The distributional approach to parts of speech

- Define parts of speech by what they can do (like chess pieces).
  - Nouns can become plural by adding –s/-es
  - Nouns can be subjects and objects of verbs
    - In English, subjects come before verbs and objects come after verbs
  - Nouns can be modified by adjectives
    - In English, most adjectives go before the noun they modify
  - What a noun refers to can be made more specific by using a determiner

The distributional approach to parts of speech

- Define parts of speech by what they can do (like chess pieces).
  - Nouns can become plural by adding –s/-es
  - Nouns can be subjects and objects of verbs
    - In English, subjects come before verbs and objects come after verbs
  - Nouns can be modified by adjectives
    - In English, most adjectives go before the noun they modify
  - What a noun refers to can be made more specific by using a determiner
- This way of defining parts of speech can be operationalized in an instruction manual so that two different people will make the same decision:
  - Is “wished” a noun or a verb in “I wished for a pony”?
  - Is “idea” a noun or a verb in “I have an idea”?

English Nouns

- Can have plurals
  - Tree
  - Trees
- Can be subjects and objects of verbs
  - Trees are tall.
  - I saw trees
- Can have determiners (articles)
  - The trees
  - These trees
- Can be modified by adjectives
  - Tall trees
- Can have possessors
  - My trees
  - The student’s trees

Not English Noun

- Can have plurals
  - Tree
  - Trees
- Can have determiners (articles)
  - The tree
  - These trees
- Can be subjects and objects of verbs
  - The tree is tall.
  - I saw trees/the tree
- Can be modified by adjectives but not adverbs
  - Tall trees
  - The tall tree
- Can have possessors
  - My trees
  - The student's trees

- Can have plurals
  - attach
  - attaches (doesn't mean plural)
- Can have determiners (articles)
  - The attach
  - These attaches
- Can be subjects and objects of verbs
  - The attach is interesting.
  - I saw attaches.
- Can be modified by adjectives
  - Fast attach
  - Quickly attach
- Can have possessors
  - *My attach
  - *The student's attach

Special types of English nouns

- Pronouns
  - Do not occur with determiners -- * *the she*
- Mass nouns
  - do not have plurals -- *two informations
- Proper nouns
  - no determiners -- *the Mt. Rushmore
  - plurals under special circumstances only
    - two Kaylas, a couple of Kevins
- Gerunds
  - deverbal (derived from verbs) nouns with -ing

There are many exceptions

- For example, even in a well-studied language like English:
  - There is mixed evidence about whether “many” is an adjective or a determiner.
  - There is mixed evidence about whether “like”, “worth”, and “near” are adjectives or prepositions.
  - “Given these problems, we should adopt this strategy.” In this sentence, “given” is starting to look like a preposition.
  - English has prepositions, not postpositions, but “ago” as in “three days ago” looks like a postposition.

Parts of speech are not discrete categories

- It is not always possible to say definitively, X is a noun or X is not a noun, for some word X.
- But parts of speech have strong prototypes: the prototypical noun, verb, preposition, adjective, etc. are easy to identify.
- So, for many NLP applications, we can pretend that parts of speech are discrete categories.

Tokenization and part-of-speech tagging

Input: raw text

```
Dr. Smith said tokenization of English is “harder than you’ve thought.”
When in New York, he paid $12.00 a day for lunch and wondered what it would
be like to work for AT&T or Google, Inc.
```

Output from Stanford Parser: http://nlp.stanford.edu:8080/parser/index.jsp
with part-of-speech tags:

```
Dr./NNP Smith/NNP said/VBD tokenization/NN of/IN English/NNP
is/VBZ ``/`` harder/JJR than/IN you/PRP 've/VBP thought/VBN ./.
''/’’

When/WRB in/IN New/NNP York/NNP ,/, he/PRP paid/VBD $/$ 12.00/CD
a/DT day/NN for/IN lunch/NN and/CC wondered/VBD what/WP it/PRP
would/MD be/VB like/JJ to/TO work/VB for/IN AT&T/NNP or/CC
Google/NNP ,/, Inc./NNP ./.
```

Part of speech tags used in the Penn Treebank

https://www.cis.upenn.edu/~treebank/

- Coordinating conjunction
- Cardinal number
- Determiner
- Existential-*there*
- Foreign word
- Preposition/subordinating conjunction
- Adjective
- Comparative adjective
- Superlative adjective
- List item marker
- Modal

Part of speech tags used in the Penn Treebank

- Singular noun or mass noun
- Plural noun
- Singular proper noun
- Plural proper noun
- Predeterminer
- Possessive ending
- Personal pronoun
- Possessive pronoun
- Adverb
- Comparative adverb
- Superlative adverb
- Particle

Part of speech tags used in the Penn Treebank

- Symbol
- *To*
- Interjection
- Base form verb
- Past tense verb
- Gerund or present participle verb
- Past participle verb
- Verb not 3rd person singular present
- Verb 3rd singular present
- Wh-determiner
- Wh-pronoun
- Possessive wh-pronoun
- Wh-adverb

The Google Universal Tag Set

- Petrov et al. 2012
- Noun
- Verb
- Adjective
- Adverb
- Pronoun
- Determiner
- Adposition
- Conjunction
- Number
- **Particle**
- Punctuation
- **Other**

These were optimized for the task of cross-lingual transfer of part-of-speech tags.

Part-of-speech taggers are typically trained on a large amount of annotated data.

Each time you need to build a part-of-speech tagger for a new language, you need to collect and annotate a lot of data.

Instead, you can transfer the parameters of the part-of-speech tagger from a related language or a typologically similar one.

Are parts of speech universal? some interesting cases

- Language doesn’t distinguish nouns and verbs
- Language doesn’t distinguish nouns and adjectives
- Language doesn’t distinguish verbs and adjectives
- Same concept expressed by different parts of speech in different languages

Hmong adjectives and verbs

“Adjective”

- *Nws yog ib tug neeg phem*
  3SG be one CLF human evil
  ‘She is an evil person.’

- *Neeg phem heev.*
  human evil very
  ‘People are very evil.’

- *Neeg phem dua cuam.*
  human evil surpass ape
  ‘People are more evil than apes.’

“(Transitive) Verb”

- *Nws yog ib tug neeg paub*
  3SG be one CLF human know
  ‘She is a person who knows’

- *Neeg paub heev.*
  human know very
  ‘People know a lot.’

- *Neeg paub dua cuam.*
  human know surpass ape
  ‘People know more than apes.’

Hmong adjectives and verbs

Note: English and Hmong have different morphosyntactic strategies for expressing comparison.

“Adjective”

- *Nws yog ib tug neeg phem*
  3SG be one CLF human evil
  ‘She is an evil person.’

- *Neeg phem heev.*
  human evil very
  ‘People are very evil.’

- *Neeg phem dua cuam.*
  human evil surpass ape
  ‘People are more evil than apes.’

This is a verb.

There is no corresponding word in Hmong.

This is a preposition.

“(Transitive) Verb”

- *Nws yog ib tug neeg paub*
  3SG be one CLF human know
  ‘She is a person who knows’

- *Neeg paub heev.*
  human know very
  ‘People know a lot.’

- *Neeg paub dua cuam.*
  human know surpass ape
  ‘People know more than apes.’

Hausa adjectives and nouns

- Schachter, page 15

mutum mai alheri

person having kindness

“a kind person.”

mutum mai doki

person having horse

“a person having a horse”

What you should remember about parts of speech

- Parts of speech are categories of words that linguists use to describe grammar.
- Parts of speech are defined by their distribution and behavior, not by their meaning.
- Parts of speech are used in many NLP applications such as Named Entity Recognition. Many NLP applications start with part of speech tagging as a preliminary step.
- Parts of speech are not discrete categories, but we can pretend that they are.
- Parts of speech are not universal, but we can pretend that they are.
- There is no correct number of parts of speech, but for the purpose of cross-lingual learning, we can pretend that there are around 12 important ones.

Part 2: Morphosyntax

- Why talk about morphology and syntax together?
- Morphosyntax is about communicative functions that can be accomplished with morphology or with syntax or both.

Interlinear Glossed Text

- Read this document:
- https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

Interlinear glossed text

(7) Mereka di Jakarta sekarang.
they in Jakarta now
‘They are in Jakarta now.’

Interlinear glossed text

Mereka	di	Jakarta	sekarang.
they	in	Jakarta	now

‘They are in Jakarta now.’

Object language (Indonesian)

metalanguage (English)

Interlinear glossed text

Mereka	di	Jakarta	sekarang.
they	in	Jakarta	now

‘They are in Jakarta now.’

Object language (Indonesian)

metalanguage (English)

The second line is the gloss. It is “interlinear” because it is between the first line and the third line.

The gloss is in the metalanguage.

The gloss **MUST** be tabbed to line up with the object language. There are many LaTeX macros for glossing.

Interlinear glossed text

Mereka	di	Jakarta	sekarang.
they	in	Jakarta	now

‘They are in Jakarta now.’

Object language (Indonesian)

metalanguage (English)

The gloss conveys information about the morophosyntax of the object language.

The third line does not convey information about the target language. It only conveys information about the metalanguage.

This is a wrong conclusion: Indonesian “di” means “are in”.

Interlinear glossed text

Mereka	di	Jakarta	sekarang.
they	in	Jakarta	now

‘They are in Jakarta now.’

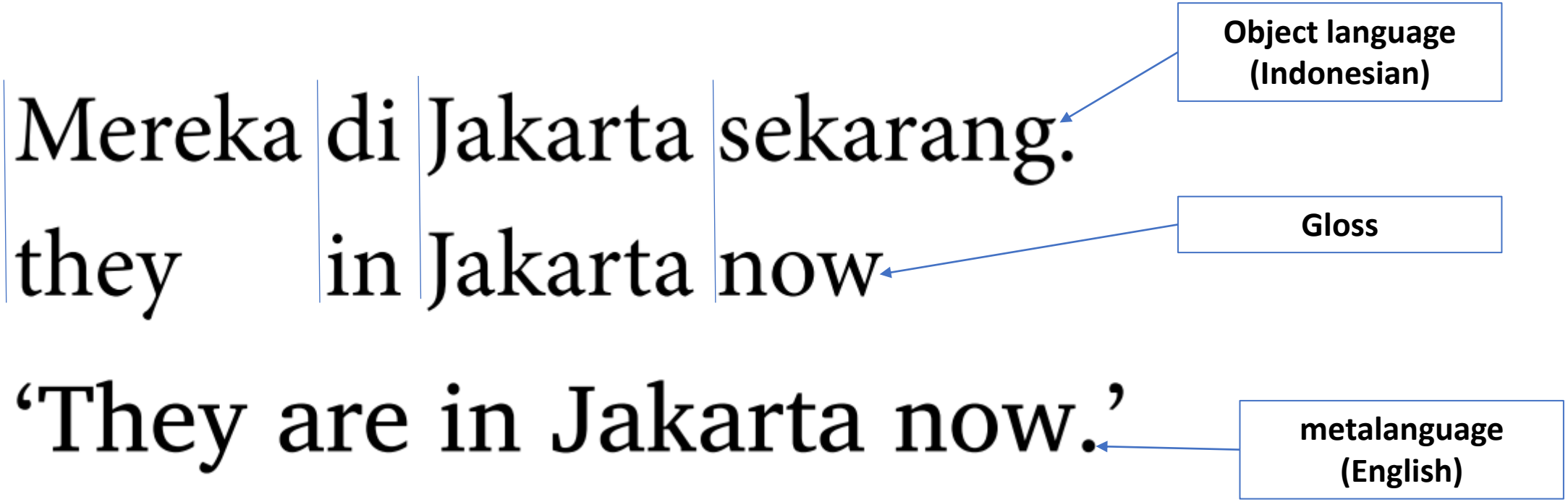
Object language (Indonesian)

metalanguage (English)

However, you can compare lines two and three to draw a conclusion about the use of different *morphosyntactic strategies* in English and Indonesian.

A correct observation: When expressing that people are located somewhere, English uses a *copula* (be-verb) strategy and Indonesian uses a *zero copula* strategy.

Interlinear glossed text



English construction:
- **Meaning: someone is located somewhere**
- **Form: pronoun copula prepositional phrase**
- **Strategy: copula**

Indonesian construction:
- **Meaning: someone is located somewhere**
- **Form: pronoun prepositional phrase**
- **Strategy: zero copula**

Mereka di Jakarta sekarang.

3.PL LOC Jakarta now

‘They are in Jakarta now.’

Morpheme labels.

It is a little misleading to say that Indonesian “di” means English “in”.

It is a little less misleading to say that Indonesian “di” is a locative preposition.

3.pl means “third person plural” the dot indicates that 3 and pl are two different meanings expressed in the same morpheme.

Russian

My	s	Marko	poexa-l-i	avtobus-om	v	Peredelkino.
1.PL	COM	Marko	go-PST-PL	bus-INS	ALL	Peredelkino

‘Marko and I went to Peredelkino by bus.’

Put a space before each morpheme and tab each morpheme with its gloss. It is easier for us to grade.

My	s	Marko	poexa	-l	-i	avtobus	-om	v	Peredelkino.
1.PL	COM	Marko	go	-PST	-PL	bus	-INS	ALL	Peredelkino

‘Marko and I went to Peredelkino by bus.’

My	s	Marko	poexa	-l	-i	avtobus	-om	v	Peredelkino.
1.PL	COM	Marko	go	-PST	-PL	bus	-INS	ALL	Peredelkino

‘Marko and I went to Peredelkino by bus.’

1	first person
ALL	allative (going toward)
COM	comitative (doing something together)
INS	instrumental (using an instrument)
PL	plural
PST	past tense

English uses the same word for both of these (“I went **with** Marko” and “I ate **with** a fork”), but not all languages use the same word/morpheme for instrumental and comitative.

My	s	Marko	poexa	-l	-i	avtobus	-om	v	Peredelkino.
1.PL	COM	Marko	go	-PST	-PL	bus	-INS	ALL	Peredelkino

1	first person
ALL	allative (going toward)
COM	comitative (doing something together)
INS	instrumental (using an instrument)
PL	plural
PST	past tense

‘Marko and I went to Peredelkino by bus.’

English Constructions:

Meaning: People do something together
Form: subject Verb-phrase [PP with-noun phrase]
Example: I went to NY with Marko.
Strategy: prepositional phrase sharing preposition with instrumental

Meaning: People do something together
Form: NP and NP VP
Example: Marko and I went to NY.
Strategy: coordinate structure

Meaning: Use a form of transportation
Form: subject verb-phrase [PP by-noun phrase]
Example: I went by bus.
Strategy: prepositional phrase sharing preposition with means/proximity (I did it by shouting. I sat by the tree.)

Russian Constructions:

Meaning: People do something together
Form: 1.pl s NP
Example: My s Marko poexali.
Strategy: plural subject plus comitative PP

Meaning: Use a form of transportation
Form: subject verb-phrase NP(instrumental case)
Example: My poexali avtobusom.
Strategy: Instrumental case marking on the form of transportation

My	s	Marko	poexa	-l	-i	avtobus	-om	v	Peredelkino.
1.PL	COM	Marko	go	-PST	-PL	bus	-INS	ALL	Peredelkino

1	first person
ALL	allative (going toward)
COM	comitative (doing something together)
INS	instrumental (using an instrument)
PL	plural
PST	past tense

‘Marko and I went to Peredelkino by bus.’

English Constructions:

Meaning: People do something together
Form: subject Verb-phrase [PP with-noun phrase]
Example: I went to NY with Marko.
Strategy: prepositional phrase sharing preposition with instrumental

Meaning: People do something together
Form: NP and NP VP
Example: Marko and I went to NY.
Strategy: coordinate structure

Meaning: Use a form of transportation
Form: subject verb-phrase [PP by-noun phrase]
Example: I went by bus.
Strategy: prepositional phrase sharing preposition with means/proximity (I did it by shouting. I sat by the tree.)

Russian Constructions:

Meaning: People do something together
Form: 1.pl s NP
Example: My s Marko poexali.
Strategy: plural subject plus comitative PP

Meaning: Use a form of transportation
Form: subject verb-phrase NP(instrumental case)
Example: My poexali avtobusom.
Strategy: Instrumental case marking on the form of transportation

The trick is that you use the third line to tell you what meanings are present because the third line is in a language you speak. So you need to learn what meanings to look for because it is not made explicit in the example. It is usually made explicit in the text surrounding the example.

Then you use the second line to understand what morphosyntactic strategies are used for expressing those meanings in the object language.

a. The North Wind and the Sun were argu -ing
DEF North Wind and DEF Sun COP.PST.3.PL argue -PROG
'The North Wind and the Sun were arguing'

b. about which one of them was strong -er,
about which one of 3.PL COP.3.SG strong -COMPAR
'about which one of them was stronger,'

c. when a traveler came by
when INDF traveler come.PST by
'when a traveler came by'

d. wear -ing a heavy coat.
wear -PROG INDF heavy coat
'wearing a heavy coat.'

e. They agree -d
3.PL agree -PST
'They agreed'

f. that whoever got the traveler to take off
COMP REL.INDF get.PST DEF traveler COMP take off
his coat first
3.SG.GEN coat first
'that whoever got the traveler to take off his coat first'

g. would be consider -ed strong -er
FUT.HYPOTH COP.INF consider -PASS strong -COMPAR
'would be considered stronger'

Example of glossing English

Form and Meaning

- **Lets start with some meanings that are so basic, you probably take them for granted.**
- Agent acts on patient:


- Figure moves against ground:



- Figure is located on ground:


- Something causes an emotion:


- Someone has something:


- Something exists:

- Someone has a role

- Someone has an identity

- Something has a property


Across languages: three morphosyntactic strategies for agent acting on patient:

How do you know whether the dog kissed the cats or the cats kissed the dog?

Word Order (English)

The cats kissed the dog.

The dog kissed the cats.

Case Marking (Turkish)

Adam –ı getir

Man -acc bring

`Bring the man' (not `The man brings something')

Agreement (Italian)

I gatt-i hanno baciato il cane

The.PL cat -PL have.3PL kiss-PART the.SG dog.SG

`The cats kissed the dog' (not `The dog kissed the cats')

This is **number**, not case. It says that there is more than one cat.

Use correct terminology:

Case markers are on nouns. The noun announces its role: I'm the agent, or I'm the patient.

Agreement markers are on verbs. The verb announces: the agent is the plural one.

Across languages: copula and zero copula for predicate nominals

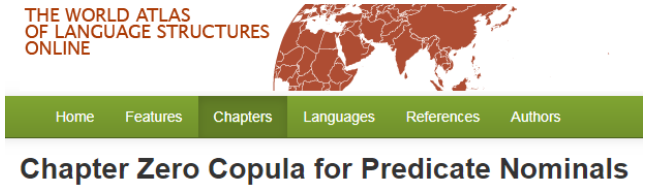
English: Obligatory copula

(1)

a. *John is a sailor.*

b. **John a sailor.*

* means “ungrammatical”. If you think of a language as a possibly infinite set of sentences, * means that this sentence is not a member of the language.



Across languages: copula and zero copula for predicate nominals

English: Obligatory copula

(1)

a. *John is a sailor.*

b. **John a sailor.*

What is a predicate nominal?

A noun like “soldier” that is predicated of another noun like “John”.

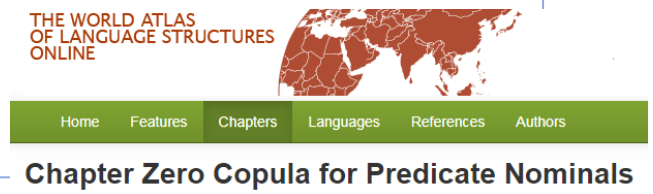
We are saying something about John (predicating something of John).

What we are saying is that he is a soldier.

What about this sentence: The soldier is John. In this sentence, we are saying something about the soldier. We are saying that it is John. In this sentence, “John” is the predicate nominal.

These sentences to not have predicate nominals:

Chris is tall (adjective)
Chris is in the room (prepositional phrase)
Chris is singing (verb)



Zero copula with predicate nominal

(6) Sinhalese (Gair 1970: 145)

unnaehee hungak presidde kene-k

he very famous person-NOM

'He is/ was a very famous person.'

(7) Tubu (Lukas 1953: 170)

sígən *lífi*

3SG.EMPH orphan

'He is/was an orphan.'

(8) Pitjantjatjara (Douglas 1959: 55)

wait *nglayayala*

man doctor

'The man is/ was a doctor.'

(9) Asmat (Voorhoeve 1965a: 168)

no *ów* *akát*

1SG man handsome

'I am/was a handsome man.'

Zero copula in the present tense. Copula in the past tense.

(10) Russian (Maria Koptjevskaja-Tamm p.c.)

a. *ona vrač*

she doctor

'She is a doctor.'

b. *on byl učenik-om*

he be.M.PST pupil-INSTR

'He was a pupil.'

(11) Maltese (Albert Borg, p.c.)

a. *Albert tabib*

Albert doctor

'Albert is a doctor.'

b. *Albert kien tabib*

Albert be.3SG.M.PST doctor

'Albert was a doctor.'

Zero copula in the present tense with third person subject. Copula with first person subject.

(12) Hungarian (Ginter and Tarnói 1991: 76, 78)

a. *én tanár vagyok*

1SG teacher be.1SG.PRES

'I am a teacher.'

b. *ő diák*

3SG.M pupil

'He is a pupil.'

Chapter Zero Copula for Predicate Nominals

Pronominal copula

(13) Hebrew (Li and Thompson 1977: 428)

Moše hu student

Moshe 3SG.M student

'Moshe is a student.'

(14) Motu (Lister-Turner and Clark 1930: 54)

lau na tau ia be hahine

I DEM/COP man she DEM/COP woman

'I am a man, she is a woman.'

(15) Turkish (Lewis 1967: 98)

Türk-üm

Turk-1SG

'I am a Turk.'

(16) Beja (Tucker and Bryan 1966: 543)

barú:k ha'á-bwa

2SG.M sheik-2SG.M

'You are a sheik.'

I actually think these are agreement markers, not pronominal copulas.

Other copulas that are not verbs

In addition to pronouns, quite a few other nonverbal items are attested in copular function. Such **particle copulas** have their origin in a variety of markers of discourse-oriented phenomena such as topicalization, backgrounding, or contrastive focus for subjects or predicates. An example of such a particle copula has been given in sentence (4), where the Kabyle copula *d'* can be identified as a general conjunctional/contrastive marker; for one thing, it is used as the conjunction 'and/with' between noun phrases. Another example is from Awtuw (Lower Sepik-Ramu; Papua New Guinea), where the item *po* is a general focus marker. Unlike in verbal sentences, the item is obligatory with nominal predicates. A similar situation holds in Vai (Mande; Liberia).

(17) Awtuw (Feldman 1986: 60, 138, 148)

a. *Yen w-æy-re*

2SG NONFACT-go-FUT

'You will go.'

b. *wan po 'w-æy-rere*

1SG FOC NONFACT-go-DESID

'I am the one who wants to go.'

c. *wan po rameyæn*

1SG FOC/COP human being

'I am a human being.'

(18) Vai (Welmers 1976: 116; Koelle 1854: 92)

a. *mànjáā mù ná'à à fɛ'ɛ*

chief FOC I him see.PERF

'It was the chief whom I saw.'

b. *mō kóro mú nda*

man old FOC/COP I

'I am an old man.'

THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

Home Features Chapters Languages References Authors

Chapter Zero Copula for Predicate Nominals

Comparison

- https://wals.info/chapter/121

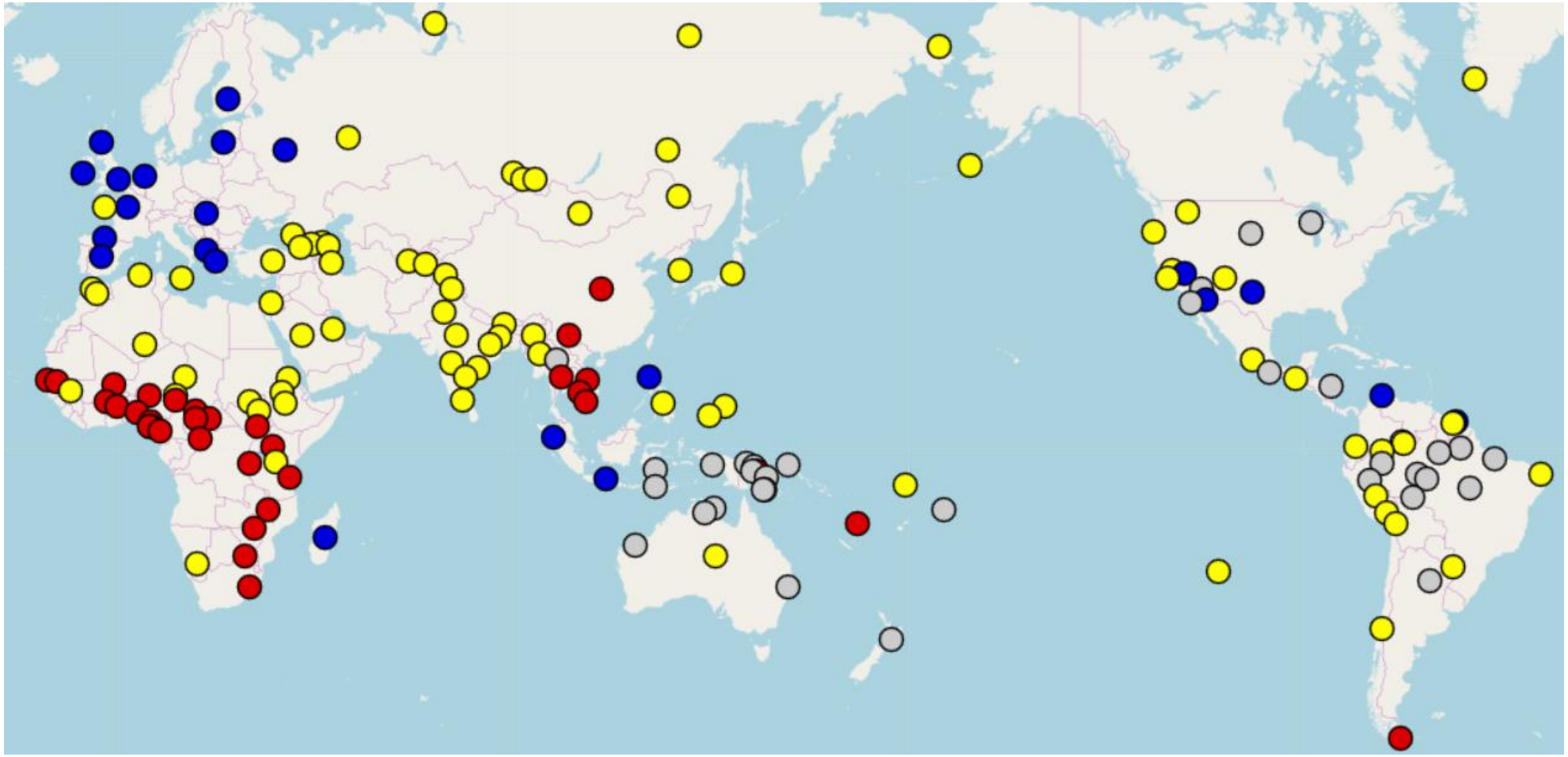
THE WORLD ATLAS OF LANGUAGE STRUCTURES ONLINE

Home | Features | Chapters | Languages | References | Authors

Feature 121A: Comparative Constructions

Values

Locational	78
Exceed	33
Conjoined	34
Particle	22



What are you comparing?

- Chris is taller than Alex.
  - Scale: height
  - Item: Chris
  - Standard: Alex

- On a **scale** of height, using Alex as the **standard** to compare against, Chris is higher on the scale.

The grammar of scalar predicative comparison in English

- Copula sentence
  - Subject BE { Adj-er | more Adj } than Y

Convention and circumlocutions

- Conventional and *grammaticalized* in English
  - X is taller than Y
- Conventional in English
  - X is tall compared to Y
- Circumlocution in English
  - X exceeds Y in height
- Extreme circumlocution in English
  - As for height, X is exceptional but I find Y lacking.

Using “exceed” for comparison

(4) Duala (Ittmann 1939: 187)

nin ndabo e kolo buka nine

this house it big exceed that

‘This house is bigger than that.’

(5) Thai (Warotamasikkhadit 1972: 71)

kăw sŭuŋ kwă kon túk kon

he tall exceed man each man

‘He is taller than anyone.’

(6) Mundari (Hoffmann 1903: 110)

sadom-ete *hati* *maranga-e*

horse-from elephant big-3SG.PRES

'The elephant is bigger than the horse.'

(7) Uzbek (Sjoberg 1963: 142)

ɔtam *u* *ɔdam-dan* *yɔš*

father.my that man-from young

'My father is younger than that man.'

(8) Estonian (Oinas 1966: 140)

kevad *on* *sügis-est* *ilusam*

spring is fall-from more.beautiful

'The spring is more beautiful than the fall.'

(9) Siuslaw (Frachtenberg 1922a: 555)

sea *his* *na-tc*

he good me-to

'He is better than me'

(10) Maasai (Tucker and Mpaayei 1955: 93)

sapuk *olkondi* *to* *lkibulekeny*

big hartebeest to waterbuck

'The hartebeest is bigger than the waterbuck.'

(11) Ahaggar Tuareg (Hanoteau 1896: 52)

kemmou tehousid *foull* *oult* *ma m*

you pretty.2SG.F upon sister of you

'You are prettier than your sister.'

(12) Tubu (Lukas 1953: 45)

sa-umma *gere* *du* *mado*

eye-his blood on red

'His eye is redder than blood.'

Using location adpositions (e.g., from, to, on) for comparison

Special comparative particle (like than)

(16) French (Bernard Bichakjian, p.c.)

tu es plus jolie que ta sœur

you are more pretty than your sister

'You are prettier than your sister.'

(17) Hungarian (Edith Moravcsik, p.c.)

István magasa-bb mint Peter

István.NOM tall-more than Peter.NOM

'István is taller than Peter.'

Conjoined

(13) Amele (Roberts 1987: 135)

jo i ben jo eu nag

house this big house that small

'This house is bigger than that house.'

(14) Menomini (Bloomfield 1962: 506)

Tata'hkes-ew nenah the kan

strong-3SG I and not

'He is stronger than me.'

(15) Malay (Lewis 1968: 157)

kayu batu bĕrat batu

wood stone heavy stone

'Stone is heavier than wood.'

Exercise

- We will identify morphosyntactic strategies in Assyrian, Swahili, Hindi, and Yoruba
  - Copula
  - Comparative
  - Copula
  - Coordination
  - Correlative: (The harder, the tighter)
- We will practice making interlinear glosses

English

- The North Wind and the Sun were arguing about which one of them was stronger, when a traveler came by wearing a heavy coat.
- They agreed that whoever got the traveler to take off his coat first would be considered stronger.
- The North Wind blew as hard as he could, but the harder he blew, the tighter the traveler wrapped his coat around him, and finally the North Wind had to give up.
- Then the sun began to shine, and the traveler immediately took off his coat.
- And so the North Wind had to admit that the Sun was stronger.

http://www.aesoplanguagebank.com/regions.html

Swahili

- Kaskazini Upepo na jua wali kuwa wana shindana gani iko na nguvuu kushinda mwingine, msafiri aka kuja na alikuwa anavaa koti mzito.
- Wali kubaliana mtu ya kwanza kutoa koti ya msafiri ndio akona nguvu kushinda ingine.
- Upepo ya kaskazini ika jaribu kupiga upepo yake yote, lakini akaona vigumu yake inapiga, zaidi msafiri anafunga koti yake karibu naye, mpaka upepo ya kaskazini ikajishinda.
- Jua ikaanza ku ngua, mpaka msafiri akatoa koti yake mara moja. Sasa Upepo ya Kaskazini ika kubali jua ikona nguvu kuishinda.

http://www.aesoplanguagebank.com/regions.html

Hindi (Romanized)

- uttarī vāyu aur sūraj bahas kar rahe the ki donõ mẽ se zyādā balavān kaun hai, jab ek rāhī garm labāde mẽ lipaṭā huā se guzarā.
- ūn donõ ne tay kiyā ki jo us rāhī kā labādā pahale utaravāne mẽ saphal hogā vo zyādā balavān hogā.
- vāyu ne tīvra gati se uṛanā śuru kiyā par vo jitanā tez hotī, utanā hī vo rāhī apane āp ko labāde mẽ kas kar lapeṭ letā.
- ākhirkār vāyu ne hār mān lī, ab sūraj bahut tez camakā aur rāhī se turant hī labādā utār diyā.
- ant mẽ uttarī vāyu ko mānanā hī paṛā ki sūraj un donõ se zyādā balavān hai.

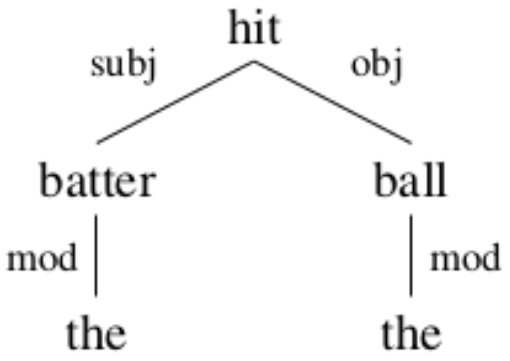
http://www.aesoplanguagebank.com/regions.html

Part 3: Syntax

Two kinds of syntactic structure

- Constituent structure
  - Tests for constituents
- Dependency Structure
  - What is a head?

Dependency



Dependency

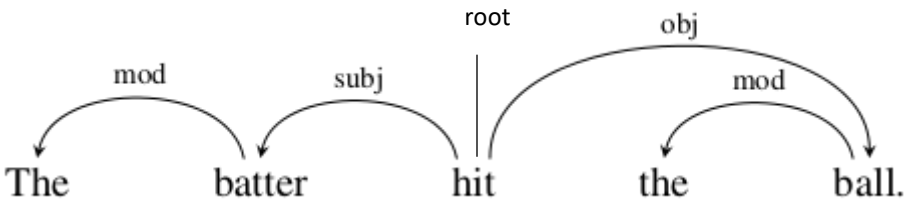
subj(2, 3)
mod(1,2)
obj(5,3)
mod(4,5)
root(3,0)

Phrase/Constituent Structure

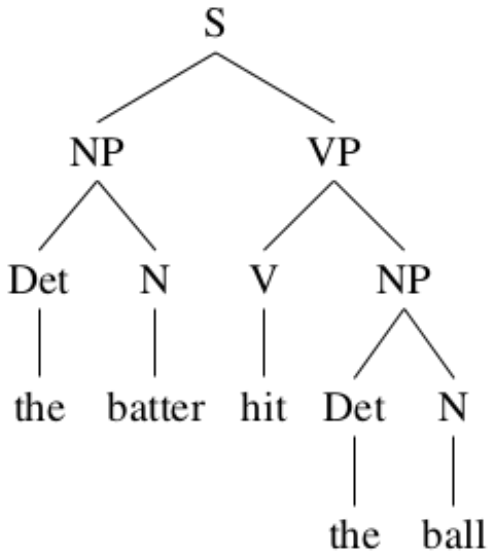
(S (NP (DET the)
(N batter))
(VP (V hit)
(NP (DET the)
(N ball))))

The batter hit the ball.

Dependency



Phrase/Constituent Structure



Constituent

- A constituent is a string of words such that there is one node that dominates those words and no other words.
- A constituent is a complete subtree consisting of one root and all the things it dominates.

Why does grammar exist?

- The most important thing in communication is to convey ***meaning***.
- So why does grammar (structure) matter and why does it even exist?

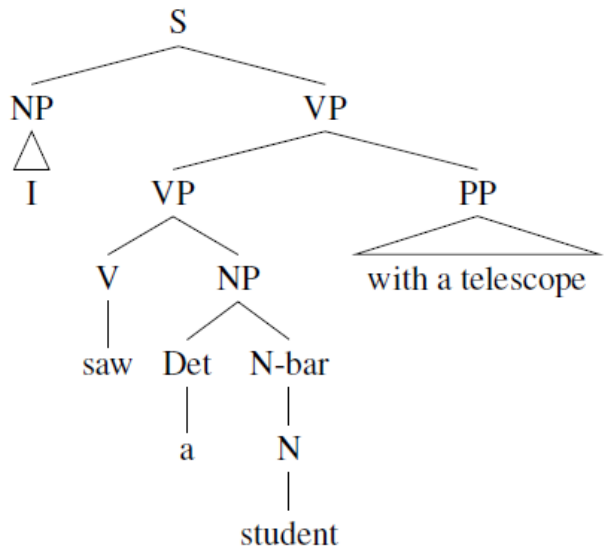
What is grammar for?

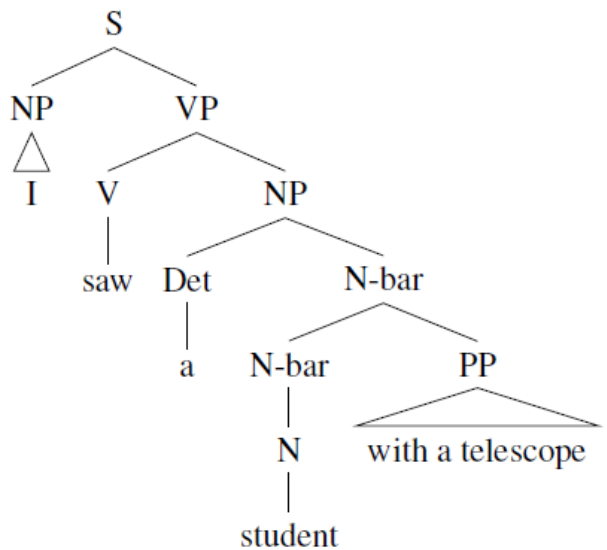
- [send] [a text] [to mom]

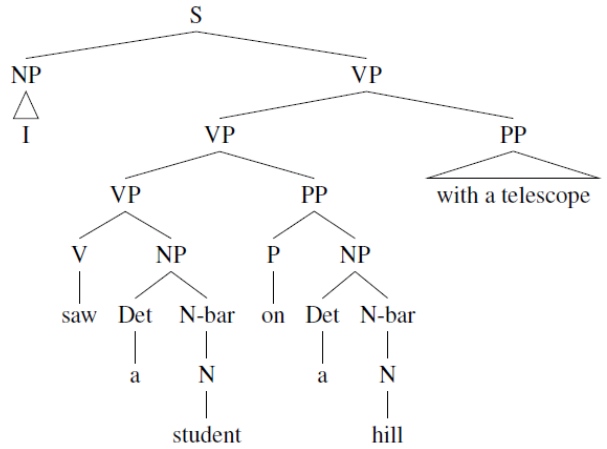
- [The old man] [is yawning]
- [The old] [man the boats]

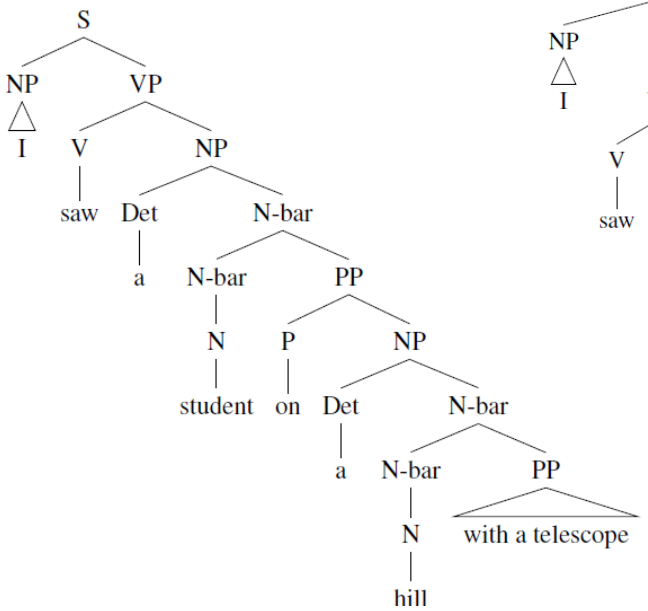
- [The cotton shirts] [are made of Egyptian cotton]
- [The cotton shirts are made of] [is soft]

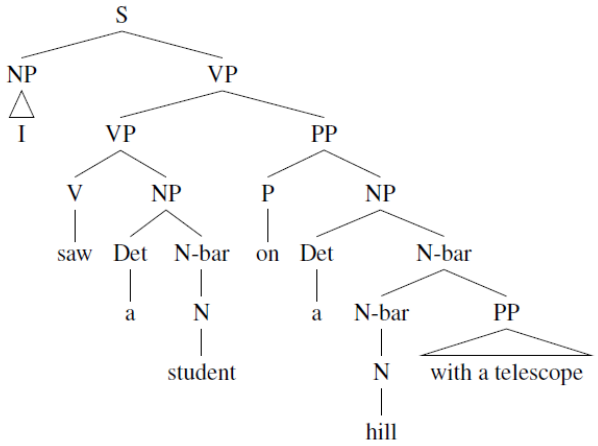
- [Time] [flies like an arrow]
- [Fruit flies] [like a banana]

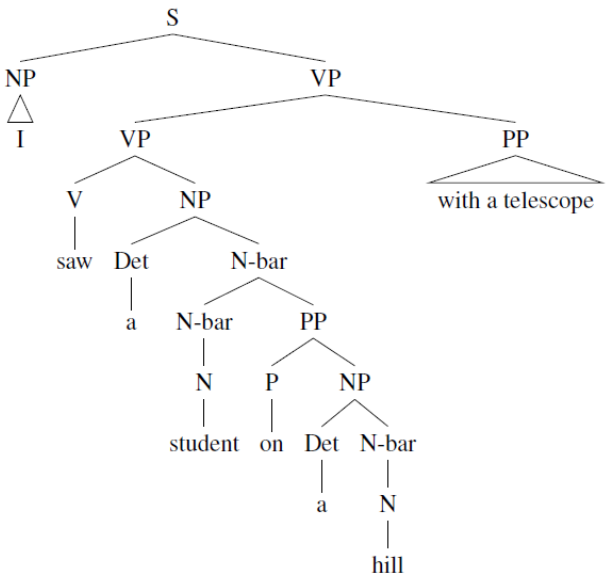

S
NP
VP
I
VP
PP
V
NP
with a telescope
saw
Det
N-bar
a
N
student

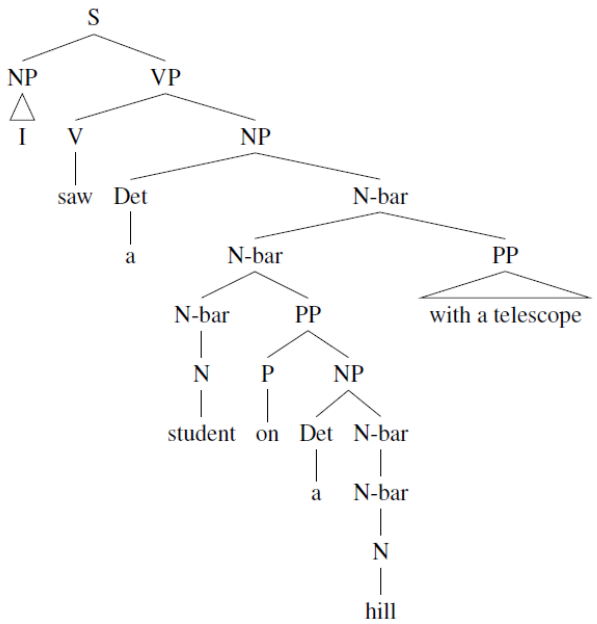

S
NP
VP
I
V
NP
saw
Det
N-bar
a
N-bar
PP
N
with a telescope
student


S
NP
VP
I
VP
PP
with a telescope
VP
PP
V
NP
P
NP
saw
Det
N-bar
on
Det
N-bar
a
N
a
N
student
hill


S
NP
VP
I
V
NP
saw
Det
N-bar
a
N-bar
PP
N
P
NP
student
on
Det
N-bar
a
N-bar
PP
N
with a telescope
hill


S
NP
VP
I
VP
PP
V
NP
P
NP
saw
Det
N-bar
on
Det
N-bar
a
N
a
N-bar
PP
student
N
with a telescope
hill


S
NP
VP
I
VP
PP
V
NP
with a telescope
saw
Det
N-bar
a
N-bar
PP
N
P
NP
student
on
Det
N-bar
a
N
hill


S
NP
VP
I
V
NP
saw
Det
N-bar
a
N-bar
PP
with a telescope
N-bar
PP
N
P
NP
student
on
Det
N-bar
a
N-bar
N
hill

Dependency Grammar

- Dependency
  - A relation between two words
  - One word is the head
  - The other is the dependent
- Each word depends on exactly one other
- Except for the root word of the tree

det(teacher-2, Every-1)
nsubj(loves-3, teacher-2)
root(ROOT-0, loves-3)
mark(teach-5, to-4)
xcomp(loves-3, teach-5)

To build a dependency tree,

- For each word, decide which word it depends on (modifies or completes its grammatical requirements)
- Then select a label such as subject, object, determiner, or modifier to name the dependency
- From experience teaching treebanking: It helps to identify the constituents before you start identifying the dependencies

Exercise

- https://arborator.ilpga.fr/q.cgi
- We will make dependency trees graphically
- The dependency trees will be automatically rendered into a data structure called CoNLL-U