

# L95: Introduction to Natural Language Syntax and Parsing

## Lecture 7

Simone Teufel

Department of Computer Science and Technology  
University of Cambridge

Michaelmas 2019/20

# Organisational

- Today: Assignment 4 open
- Today: Feedback on Assignment 2
- Today: Clause types
- Today: Dependencies

## Reading:

- Assignment 3: Submit on Monday; make it **readable**
- You read Chapters 15.1-15.3
- End of syntax after today; Paula Buttery will do 4 lectures.
- Sometime before Mid-Nov: Read Chapter 16.1-16.4 (Semantics) and section 7 in handout.
- I still haven't received any logic worksheets – really all clear?

## Context-free grammar from J&M, chapter 12

S	→	NP VP
NP	→	Pronoun   Proper-Noun   Det Nominal
Nominal	→	Nominal Noun   Noun
VP	→	Verb   Verb NP   Verb PP   Verb NP PP   Verb S
PP	→	Preposition NP
Det	→	NP 's
Nominal	→	Nominal PP
Nominal	→	Nominal GerundVP
Nominal	→	Nominal RelClause
RelClause	→	(who   that) VP

# Coordination

NP → NP and NP  
Nominal → Nominal and Nominal  
VP → VP and VP  
S → S and S  
X → X and X

## Non-declarative sentences

S → VP

S → Aux NP VP

S → Wh-NP VP

S → Wh-NP Aux NP VP

# The lexicon

Det	→	a   the   an   this   these   that
Verb	→	is   prefer   like   need   want   fly
Noun	→	flight   breeze   trip
Pronoun	→	me   I   you   it
Proper-Noun	→	Alaska   Baltimore   Los Angeles   Chicago   United
Preposition	→	from   to   on   near
Conjunction	→	and   or   but

## Feedback on Assignment 2

Some fundamentals:

- What form should the tree have?
- Base your analysis on J&M grammar; adapt the rules
- Inventing new rules
  - e.g. for adjective modification
- Inventing new labels
  - Necessary for complicated RC
- Don't forget about ambiguity and writing down all analyses
- Don't drop inconvenient parts of the sentence

## 3 kinds of rules

- Subcategorisation rules
- Modification rules
- Specification rule
- And . . . nothing much else!
- *strain resources to breaking point*



# Flat rules

- Flat rules are not a good idea
- $S \rightarrow NP VP CompS$
- They mean you mixed more than one principle into a single rule
- Think about overgeneration
- Think about undergeneration

# Difficult constructions

- Parentheticals

Nom  $\rightarrow$  Nom Parens

Parens  $\rightarrow$  ( Nom ) | - Nom - | : Nom

- Relative clauses (later)
- NPs without a determiner

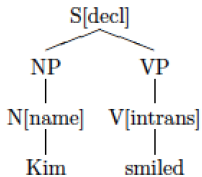
NP  $\rightarrow$  Det Nom | Nom[plural] | Nom[mass]

## Difficult constructions

- *Letters delivered on time by oldfashioned means*
- *most probable tag sequence* – noun compounding
- *that I ever liked* – what to do with the adverb?
- *the only option available*
- *at most two men*

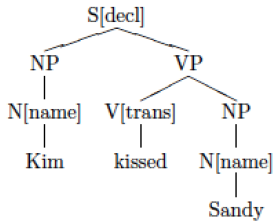
# Intransitive verb

a Kim smiled

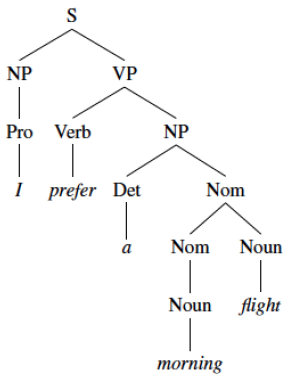


# Transitive verb

b Kim kissed Sandy

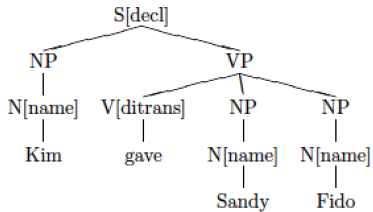


## From J and M, chapter 12.3.3



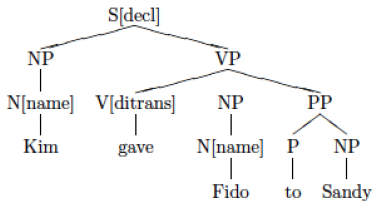
# Ditransitive verb 1

c Kim gave Sandy Fido



## Ditransitive verb 2

d Kim gave Fido to Sandy





# Types of Clauses

- subordinate clauses [finite, -ing, infinitive]
  - *I can't believe that he tweeted that*
  - *I don't like to fish in polluted rivers*
  - *I made him do the dishes*
- WH-clauses
  - *I asked who was at the party*
- relative clauses [object/subject, reduced, non-restrictive/restrictive]

# Relative clauses

- Object vs subject RC

- *the man who filmed her was fellini*
- *the man who she filmed was fellini*

Relclause\_subj -> WDT VP

Relclause\_obj -> WDT NP VP

- Reduced RC

- *the paper presented here will address. . .*
- *the director filming in studio 2 is tarantino*

- Restrictive vs non-restrictive

- *the Iranian runners who reached the goal within 2 hours were tired*
- *the Iranian runners, who reached the goal within 2 hours, were tired*

- With preposition

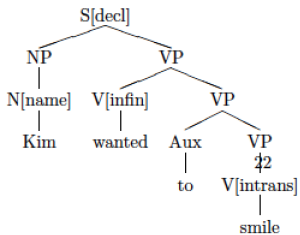
- *The person who(m) I learned most from*
- *The person from who(m) I learned most*

Rel -> WDT

Rel -> TO WDT

# Subject Control verb

e Kim wanted to smile



## Control vs. Raising Verbs

- Control: Subject or object is semantically an argument of the verb
  - *Kim tried to enjoy the party* [subject control]
  - *Kim persuaded Lee to go to Paris* [object control]
- Raising: Subject or object is semantically not an argument of the verb
  - *Kim seemed to enjoy the party.* [subject raising]
  - *Kim expects Lee to have gone to Paris.* [object raising]

## RASP Dependencies

*My aunt's can opener can open a drum*

```
(|ncsub| |open:7| |opener:5| |_)
(|aux| |open:7| |can:6|)
(|dobj| |open:7| |drum:9|)
(|det| |drum:9| |a:8|)
(|ncmod| |poss| |opener:5| |aunt:2|)
(|ncmod| |_ |opener:5| |can:4|)
(|det| |aunt:2| |My:1|)
```

All GRs are of the following form:

```
(GR-type optional-subtype head dependent optional-initial-GR)
```

# The RASP relation hierarchy

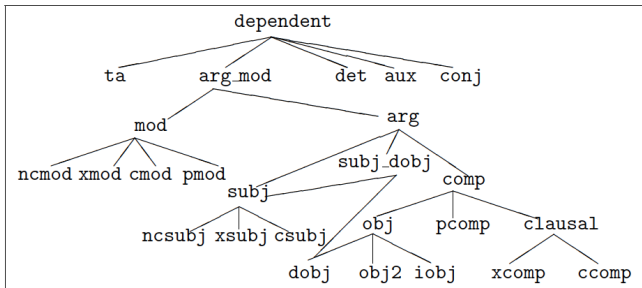


Figure 1: The GR hierarchy

# The RASP grammatical relation set (1)

conj	relation between a coordinator and the head of a conjunct.
aux	relation between main verbs as (semantic) head and auxiliary dependents.
det	relation between articles, quantifiers, partitives and other single word forms which can begin NPs and the head of the NP.
ncmod	relation between non-clausal modifiers and their heads. Subtypes: default ( ), part(itive), prt(particle), poss(essive), num(ber), ta(text adjunct), and ij(interjection).
xmod	unsaturated predicative relation between modifiers (VPs, APs) and heads. There are subtypes default ( ) and "to" for infinitive VPs
cmod	saturated relation between clausal (S) modifiers and heads. There are subtypes default ( ) and complementizer
xsubj	relation between unsaturated predicative subjects (VP, AP) and verbal heads
passive	relation naming the head of a passive VP

## The RASP grammatical relation set (2)

nsubj	relation between non-clausal subjects (NPs, PPs) and their verbal heads
xsubj	relation between unsaturated predicative subjects (VP, AP) and verbal heads
csbj	relation between saturated clausal subjects (S/V2) and verbal heads
dobj	relation between verbal or prepositional head and the head of the NP to its immediate right
obj2	relation between verbal heads and the head of the second NP in a double object construction
iobj	relation between a head and the preposition of a PP argument when the PP complement is a NP
pcomp	relation between a head and the preposition of a PP argument when the PP complement is itself a PP
xcomp	relation between a head and an unsaturated VP complement
ccomp	relation between a head and the head of a saturated clausal complement, either finite, subjunctive, headed by a wh-element or a non-finite "small clause"
ta	relation between a head and the head of a text adjunct delimited by some punctuation



## nclsubj – non-clausal subject

- nclsubj encodes binary relations between non-clausal subjects (NPs, PPs) and their verbal heads.
- There are four **initial** GR values: default/subj ( ), underlying object (obj), raising subject (rais) and inverted (inv) which is used for locative (PP, AdvP) inversion and quote inversion (*said Kim*):

*the upset man*

(nclsubj upset man obj)

(passive upset)

## nsubj – unsaturated predicative complements

- nsubj is also used for understood subjects of unsaturated predicative complements and some modifiers

*Kim wants to go*

(nsubj want Kim)

(ccomp want go)

(nsubj go Kim)

## csubj – clausal subject

- csubj is a binary relation between saturated clausal subjects (S/V2) and verbal heads.
- The subtype slot is filled by the complementizer if the clause is finite and left empty for non-finite ‘small clause’ like *her coming matters*:

*that he came matters*

(csubj matters came that)

(ncsubj came he)

## doj

doj is a binary relation between verbal or prepositional head and the head of the NP to its immediate right.

*She gave it to Kim*

(doj gave it)

(ncsubj gave She \_)

(ioj gave to)

(doj to Kim)

## obj2

- obj2 is a binary relation between verbal heads and the head of the second NP in a double object construction

*She gave Kim toys*

(obj2 gave toys)

(dobj gave Kim)

(ncsubj gave She \_)

## iobj

- iobj is a binary relation between a head and the preposition of a PP argument when the PP complement is a NP

*Kim flew to Paris from Geneva*

(ncsubj flew Kim \_)

(iobj flew to)

(iobj flew from)

(dobj to Paris)

(dobj from Geneva)

*a premium of \$30.5 million*

(det premium a)

(iobj premium of)

(dobj of \$)

(ncmod num \$ 30.5)

(ncmod num 30.5 million)

## ncmod – non-clausal modifier

- ncmod encodes binary relations between non-clausal modifiers and their heads.
- There are subtypes: default ( ), part(itive), prt(particle), poss(essive), num(ber), ta(text adjunct), and ij(interjection).
- The default case covers most pre-/post-modification.

*the old man in the barn slept*

(ncmod \_ man old)

(ncmod \_ man in)

(dobj in barn)

## Other ncmmod cases

- Numbers are identified as special types of modifier where possible.
- Possessives are treated as relations between head and dependent nouns:  
*the butcher's shop*  
(ncmod poss shop butcher)  
where the head can be ellip(tical).
- Partitive predeterminers: *all the men*:  
(ncmod part men all)
- Verbal particles: *look up the word*:  
(ncmod prt look up)



## det – articles, quantifiers, partitives and similar

- det encodes a binary relation between articles, quantifiers, partitives and other single word forms which can begin NPs and the head of the NP.

*Some men came*

(det men Some)

(ncsubj came men \_)

## xsubj – unsaturated predicative subjects

- xsubj encodes binary relations between unsaturated predicative subjects (VP, AP) and verbal heads.
- This relation only has non-default subtype value inverted (inv) used for extraposition examples like *to go appears difficult*:

*leaving matters*

(xsubj matters leaving \_)

## conj – conjuncts

- conj encodes relations between a coordinator and the head of a conjunct.
- There will be as many such binary relations as there are conjuncts of a specific coordinator.

*Kim likes oranges, apples, and satsumas or clementines*

```
(ncsubj likes Kim _)  
(dobj likes and)  
(conj and oranges)  
(conj and apples)  
(conj and or)  
(conj or satsumas)  
(conj or clementines)
```

## aux – auxiliaries

- aux encodes relations between main verbs as (semantic) head and auxiliary dependents.
- There are as many such binary relations as there are auxiliaries.
- If a copular or main verb form of an auxiliary is present then it is the head of any such aux relation.
- The head of aux can be ellip(tical) as in Kim will.

*Kim has been sleeping*

(ncsubj sleeping Kim \_)

(aux sleeping has)

(aux sleeping been)

## xmod – unsaturated predicative relations between modifiers

- xmod encodes binary unsaturated predicative relations between modifiers (VPs, APs) and heads.
- Subtype “to” is used when the modifier is an infinitive VP (though the current grammar doesn’t always recover it)

*who to talk to*

(xmod to who talk)

(iobj talk to)

(dobj to who)

## cmod

- cmod encodes binary saturated relations between clausal (S) modifiers and heads.
- There are subtypes default ( ) and complementizer 'that'

*although he came, Kim left*

(cmod \_ left although)

(ccomp although came)

## pmod

- pmod encodes binary relations between PP modifiers with PP complements and heads

*he went, off into the darkness*

(pmod went off)

(pcomp off into)

(dobj into darkness)

## pcomp

- pcomp is a binary relation between a head and the preposition of a PP argument when the PP complement is itself a PP

*Kim climbed through into the attic*

```
(ncsubj climbed Kim _)  
(pcomp climbed through)  
(pcomp through into)  
(dobj into attic)
```



## xcomp

- xcomp is a binary relation between a head and an unsaturated VP complement.
- It has subtypes: default ( ) and 'to', the latter indicating an infinitival complement.

*Hooker's philosophy was to build and sell.*

(ncsubj was philosophy \_)

(xcomp to was and)

(conj and build)

(conj and sell)

(ncmod poss philosophy Hooker)

*Kim thought of leaving*

(ncsubj thought Kim \_)

(xcomp \_ thought of)

(xcomp \_ of leaving)

## ccomp

- ccomp is a binary relation between a head and the head of a saturated clausal complement, either finite, subjunctive,
- headed by a wh-element or a non-finite 'small clause'.
- It has subtypes: default ( ) and 'that'.
- The head of the dependent clause is usually the verb but can be the subject of the 'small clause'

*Kim asked about him playing rugby*

(nsubj asked Kim \_)

(ccomp \_ asked about)

(ccomp \_ about him)

(nsubj playing him \_)

(dobj playing rugby)

## ta – text adjunct

- ta is a binary relation between a head and the head of a text adjunct delimited by some punctuation
- Subtypes: quote, brack(et), dash, colon, comma, bal(anced), end, echo, tag (questions), refl(exive), voc(ative).
- Balanced text adjuncts have matching delimiting punctuation (e.g. commas or dashes at both boundaries).
- End text adjuncts usually occur sentence-finally and thus the matching punctuation mark at the right boundary is promoted to a full stop.
- The remaining subtypes attempt to infer some of the semantic/discourse import of a comma from information in the PoS tagset concerning nominal lexical types.

*He made the discovery: Kim was the abbot; Lee was the host.*

(ncsubj made He \_)  
(dobj made discovery)  
(ta colon discovery was)

(ncsubj was Lee \_)