

Challenges of Processing South Asian Languages (CPSAL)

Kengatharaiyer **Sarveswaran**

University of Konstanz, Germany

University of Jaffna, Sri Lanka.

Tafseer Ahmed

Senior NLP Scientist

Alexa Translations

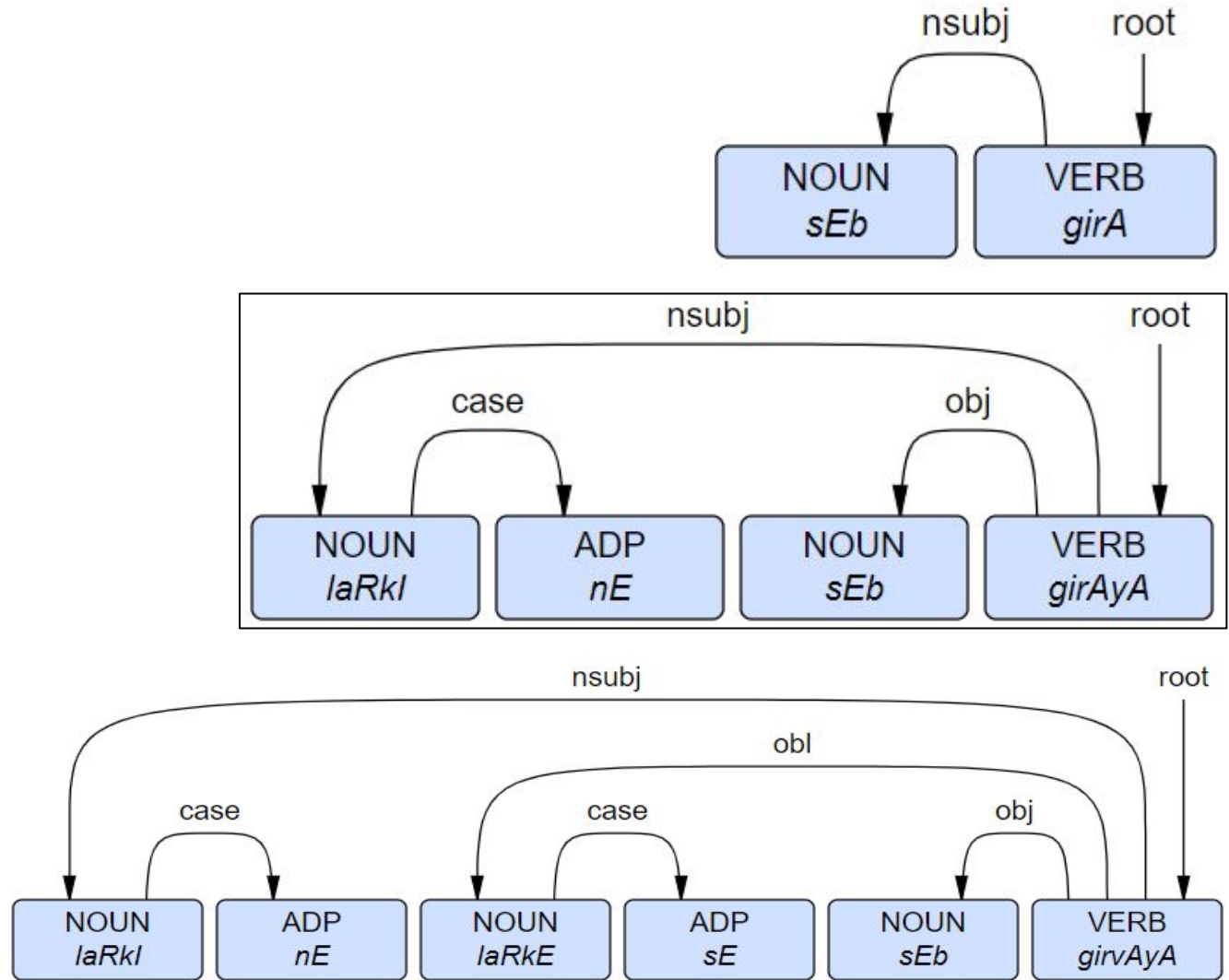
Course outline

- **Topics (Tentative):**
 - Day 01: Languages, Scripts, and Encoding of South Asian Languages.
 - Day 02: Phonology, Transliteration and Morphology of South Asian Languages.
 - Day 03: More on Morphology, Part of Speech and Multi-word tokenisation
 - **Day 04: Syntax, Morphosyntax, and Semantics of South Asian Languages.**
 - Day 05: Deep Learning for South Asian Languages and winding up the course.

Urdu - Some interesting constructions

Causatives

Causatives - Urdu



Causatives - Urdu

# sent_id = 1							
# text = sEb gir-A سیب گرا							
1	sEb	sEb	NOUN	NN	Case=Nom Gender=Masc Number=Sing Person=3	2	nsubj
2	girA	gir	VERB	VM	Aspect=Perf Gender=Masc Number=Sing VerbForm=Part V	0	root
# sent_id = 2							
# text = laRkl nE sEb gir-A-yA لڑکی نے سیب گرایا							
1	laRkl	laRkl	NOUN	NN	Case=Acc Gender=Fem Number=Sing Person=3	4	nsubj
2	nE	nE	ADP	PSP	AdpType=Post	1	case
3	sEb	sEb	NOUN	NN	Case=Nom Gender=Masc Number=Sing Person=3	4	obj
4	girAyA	girA	VERB	VM	Aspect=Perf Gender=Masc Number=Sing VerbForm=Part V	0	root

Do we need causative as a feature, and *gir* as the lemma of *girAyA* and *girvAyA* too?

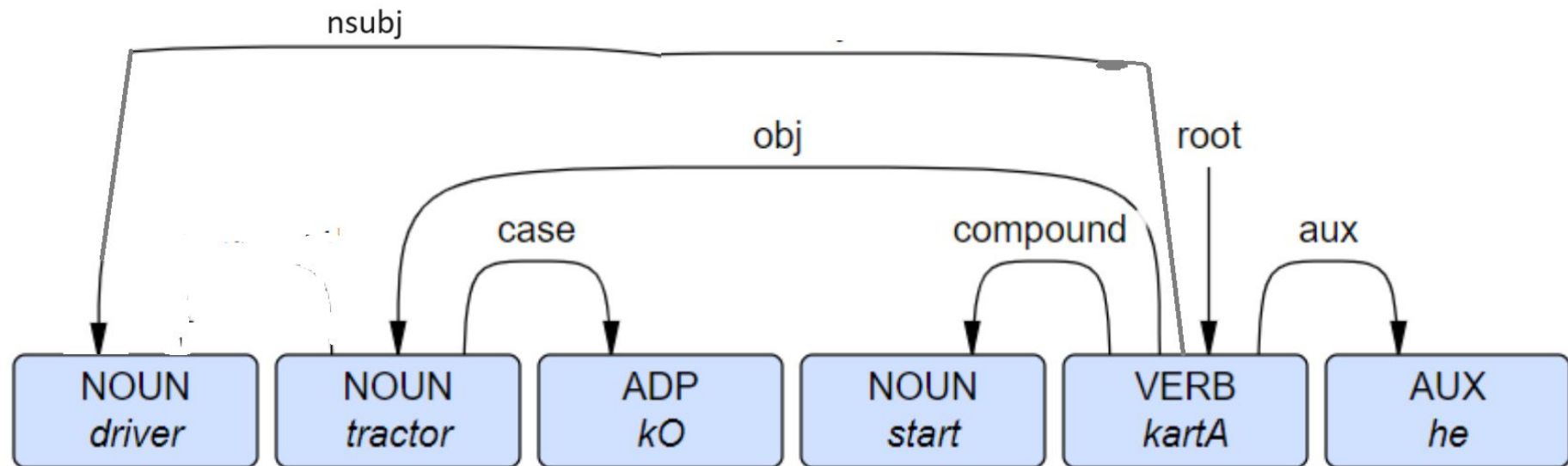
Complex predicates

Complex Predicates

ڈرائیور ٹریکٹر کو سٹارٹ کرتا ہے۔

DrAIvar TrEkTar=kO sTArT kartA hE
driver.Masc.Sg tractor.Masc.Sg=Acc start do.Imperf.Masc.Sg be.Pres.3.Sg

'The driver starts the tractor.'

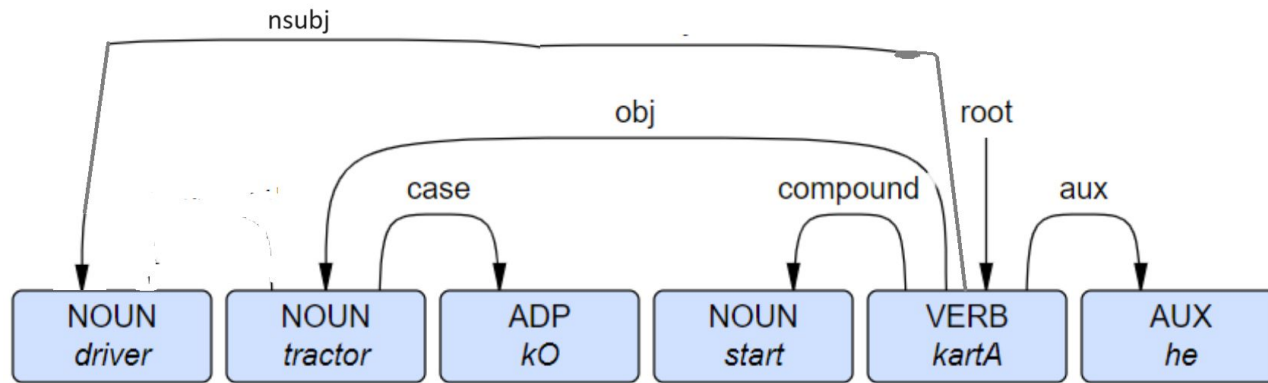


Complex Predicates

ڈرائیور ٹریکٹر کو سٹارٹ کرتا ہے۔

DrAIvar TrEkTar=kO sTArT kartA hE
driver.Masc.Sg tractor.Masc.Sg=Acc start do.Imperf.Masc.Sg be.Pres.3.Sg

'The driver starts the tractor.'

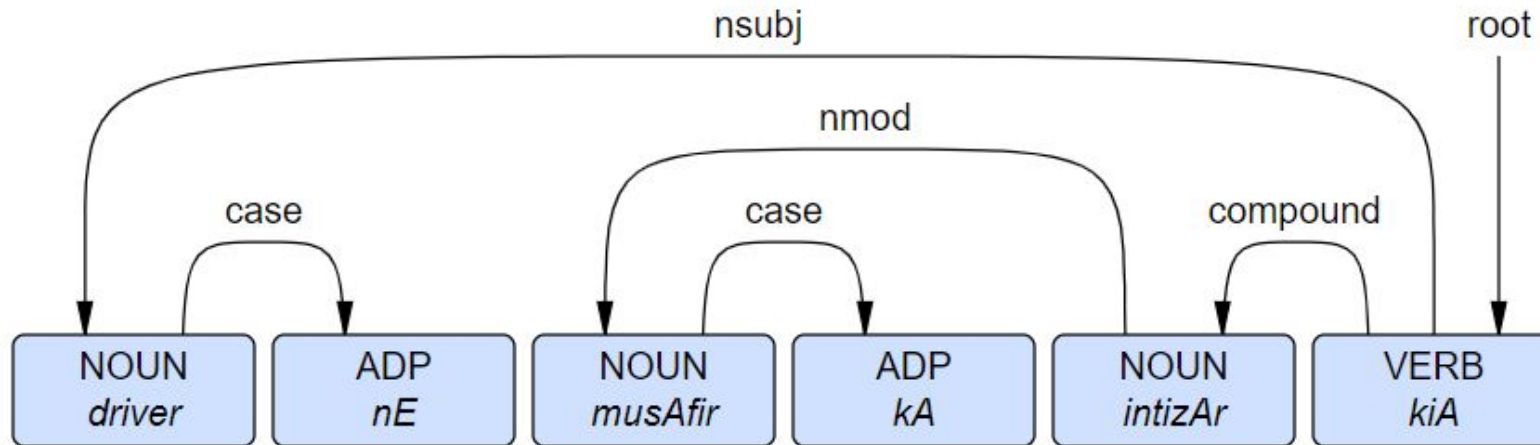


PRED	'kar<[9:DrAIvar], 'sTArT<[4:TrEkTar]>'>'
VTYP	COMPLEX-PRED-FORM kar, COMPLEX-PRED nv
TNS-ASP	TENSE pres, MOOD indicative, ASPECT impf
LEX-SEM	AGENTIVE +
OBJ	PRED 'TrEkTar' SEM-PROP 8 SPECIFIC + NTYPE NSEM 7 COMMON count NSYN common PERS 3, NUM sg, GEND masc, CASE acc
SUBJ	PRED 'DrAIvar' NTYPE NSEM 11 COMMON count NSYN common PERS 3, NUM sg, GEND masc, CASE nom
0	CLAUSE-TYPE decl

Complex Predicates

- Genitive marked argument

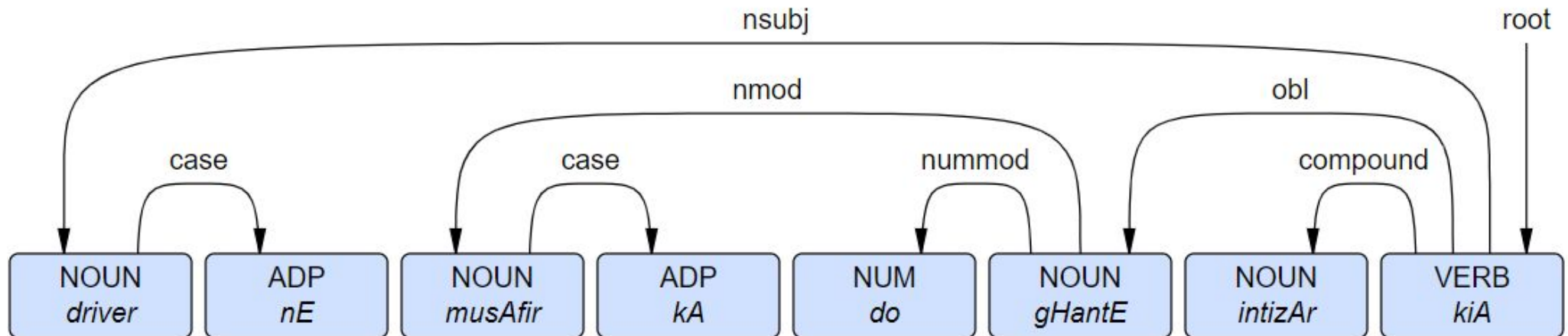
```
driver=nE    [musAfir=kA    intizAr]    kiA  
driver=Erg  passenger=Gen  wait.N      do.Perf  
'The driver waited for the passenger.'
```



Complex Predicates

- Genitive marked argument

driver=nE musAfir=kA do gHanTE intizAr kiA
driver=Erg passenger=Gen two hour.Pl wait.N do.Perf
'The driver waited for two hours for the passenger.'



Complex Predicates

biccHU=nE meNDak=sE kAI accHI bAtEN k-IN
 scorpion=Erg frog=Inst several good discussion do-Perf
 The scorpion discussed several good things with the frog.'

bAT 'thing/matter' has modifiers.

"biccHU nE meNDak sE kAI acCHI bAtEN kIN"

	[PRED	'kar<[1:biccHU], 'bAt<[35:meNDak]>'>']
			[
			PRED	'biccHU'
			CHECK	[_NMORPH obl]
SUBJ			NTYPE	[NSEM [COMMON count] [NSYN common]
			1	CASE erg, GEND masc, NUM sg, PERS 3]
				[
			PRED	'bAt'
			ADJUNCT	{ [PRED 'acch' 72[ATYPE attributive, DEGREE positive, GEND fem, NUM pl]] }
			CHECK	[_RESTRICTED +]
			LEX-SEM	[AGENTIVE +]
OBJ			NTYPE	[NSEM [COMMON count] [NSYN common]
			SPEC	[QUANT 70 [PRED 'kAI' [NUM pl]]]
			TNS-ASP	[ASPECT perf, MOOD indicative]
				[CASE nom, CLAUSE-TYPE decl, GEND fem, NUM pl, PASSIVE -, VTYPE complex-pred]
				[
			PRED	'meNDak'
			CHECK	[_NMORPH obl]
OBL			NTYPE	[NSEM [COMMON count] [NSYN common]
			35	CASE inst, GEND masc, NUM sg, PERS 3]
]

Complex Predicates

- **Issue: Associating the arguments with both elements complex predicates**
 - ParGram has implemented a solution. UD needs standardization.

N/Adj +Verb and other complex predicates:

http://www.lrec-conf.org/proceedings/lrec2012/pdf/474_Paper.pdf

Tense and Aspect

Tense and Aspect - Urdu

TNS-ASP	TENSE	R-S	precede/follow/overlap		
		DISTANCE	standard/far/close		
	ASPECT	E-R	precede/follow/overlap		
		DISTANCE	standard/far/close		
		TELIC	+/-		
		DUR	+/-		
		CONT	+/-		
		ITER	+/-		

<https://dokumen.tips/documents/tense-and-aspect-in-urdu.html>

Tense and Aspect - Urdu

vo	ter-t-A	he
Pron.3Sg	swim-imperf-M.Sg	be.Pres.M.Sg
'He swims.' (daily)		

vo	ter	rah-A	he
Pron.3Sg	swim	stay-Perf.M.Sg	be.Pres.M.Sg
'He is swimming'			

vo	ter-A
Pron.3Sg	swim-Perf.M.Sg
'He swam / He has swum.'	

Tense and Aspect - Urdu

- (35) a. vo ter-t-a hε
Pron.3.Sg.Nom swim-Impf-M.Sg be.Pres.3.Sg
'He swims (e.g., daily).'
- b. vo ter-t-a ja-t-a hε
Pron.3.Sg.Nom swim-Impf-M.Sg go-Impf-M.Sg be.Pres.3.Sg
'He keeps on swimming.'
(e.g., 20 minutes duration, 3–4 times daily)
- c. vo ter-t-a cal-a ja-t-a hε
Pron.3.Sg.Nom swim-Impf-M.Sg walk-Perf.M.Sg go-Impf-M.Sg be.Pres.3.Sg
'He keeps on swimming.'
(e.g., 1–2 hours duration, 3–4 times daily)
- d. vo ter-t-a cal-a ja rah-a hε
Pron.3.Sg.Nom swim-Impf-M.Sg walk-Perf.M.Sg go stay-Perf.M.Sg be.Pres.3.Sg
'He keeps on swimming.'
(e.g., 1–2 hours duration, ' and is still continuing)

Tense and Aspect - Urdu

- In Urdu/Hindi, aspect is determined by the sequence of lemmas and morphological forms of the aspectual auxiliary.
- See the provided paper for more sequences of tense/aspect auxiliaries.

Light verbs

(Verb +) Light Verbs - Urdu

- a. nadya=ne xat **lik^h li-ya**
Nadya.F=Erg letter.M.Nom write take-Perf.M.Sg
'Nadya wrote a letter (completely).'
- b. nadya=ne makan **bana di-ya**
Nadya.F=Erg house.M.Nom make give-Perf.M.Sg
'Nadya built a house (completely, for somebody else).'

(Verb +) Light Verbs - Urdu

- d. nadya **gir par-i**
Nadya.F.Nom fall fall-Perf.F.Sg
'Nadya fell (down, suddenly).'
- e. nadya **gir gε-yi**
Nadya.F.Nom fall go-Perf.F.Sg
'Nadya fell (down).'
- f. nadya **bol ut^h-i**
Nadya.F.Nom speak rise-Perf.F.Sg
'Nadya spoke up/broke into speech (suddenly, forcefully).'

Dative subjects

Dative Subjects - Urdu

use ye xıyal baha-ya
Pron.3.Sg.Dat this.Nom idea please-Perf.M.Sg
'This idea pleases him.'

muj^he buxar hε
Pron.1.Sg.Dat fever.M.Nom be.Pres.3.Sg
'I have a fever.'

<https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/nijmegen-hnd.pdf>

Prodrops

ProDrop

- Any or all arguments can be dropped, as both subject and object are dropped in the second sentence.

kya tum=nE nadya=kO kHana diya?
INTG You=Erg Nadya=Dat food give-Perf
'Did you give food to Nadya.'

ji, ~~main=nE~~ ~~us=kO~~ ~~kHana~~ diya?
Yes, ~~I=Erg~~ ~~she=Dat~~ ~~food~~ give-Perf
'Yes, I gave food to her.'

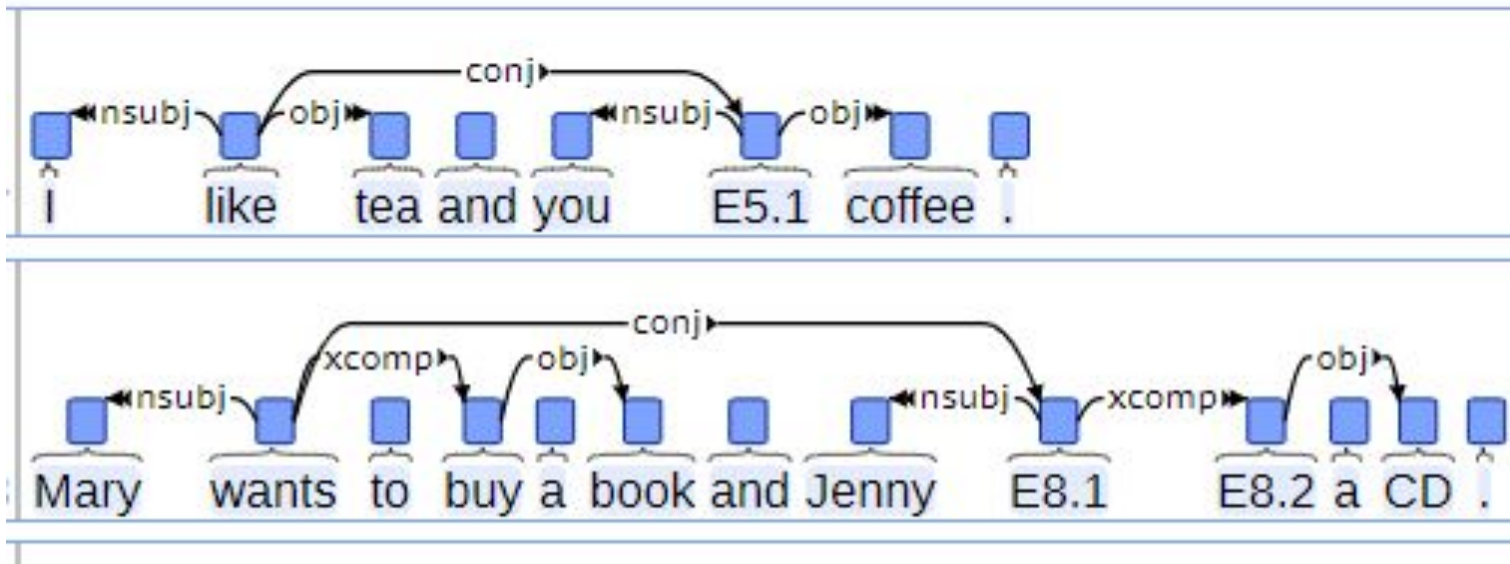
<https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/utrecht01-hnd.pdf>

Ellipsis and Enhanced Dependencies

- I like tea and you coffee.
- Mary wants to buy a book and Jenny a CD.
- They had left the company, many for good.

Ellipsis and Enhanced Dependencies

1	I	-	-	-	-	2	nsubj	2:nsubj	-
2	like	-	-	-	-	0	root	0:root	-
3	tea	-	-	-	-	2	obj	2:obj	-
4	and	-	-	-	-	5	cc	E5.1:cc	-
5	you	-	-	-	-	2	conj	E5.1:nsubj	-
5.1	-	-	-	-	-	2	conj	2:conj	-
6	coffee	-	-	-	-	5	orphan	E5.1:dobj	-



<https://universaldependencies.org/u/overview/enhanced-syntax.html>

Pro Drop with NULL element(s)

ji, main=nE us=kΘ kHana diya?
Yes, I=Erg she=Dat food give-Perf
'Yes, I gave food to her.'

1	ji	—	—	—	—	3	discourse	3:discourse
2	,	—	—	—	—	3	punct	3:punct —
2.1	—	—	—	—	—	3	nsubj	3:nsubj —
2.2	—	—	—	—	—	3	iobj	3:iobj —
2.3	—	—	—	—	—	3	obj	3:obj —
3	diya	—	—	—	—	0	root	0:root —