

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

**Tamil - Some interesting
constructions**

Tense and Aspect/Modal - Tamil

- Tense is expressed morphologically.
- Auxiliaries can be added as suffixes to the main verb to indicate aspect and modal forms.

அழிந்துக்கொண்டிருந்தேன்
Aḷintukkoṇṭiruntēn
(I) was being destroyed

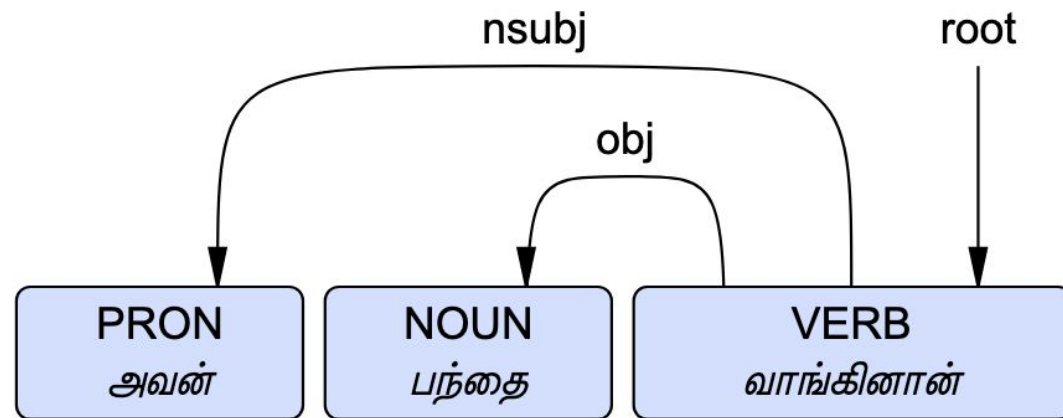
| | | | | | |
|-----------------|-------------------------------------|-----------------------------|-----------------------------|----------------------------|--|
| அழி | ந்து | கொண்டு | இரு | ந்த் | ஏன் |
| <i>aḷi</i> | <i>ntu</i> | <i>koṇṭu</i> | <i>iru</i> | <i>nt</i> | <i>ēn</i> |
| root destroy | transitivity marker intransitive | aspect marker continuous | aspect marker continuous | tense marker past tense | person marker first person, singular |

Object & Definiteness marking - Tamil

- Objects are not marked with the ACC marker (ai) if the noun is an irrational indefinite noun.
- Models may become confused because there is no overt marking.

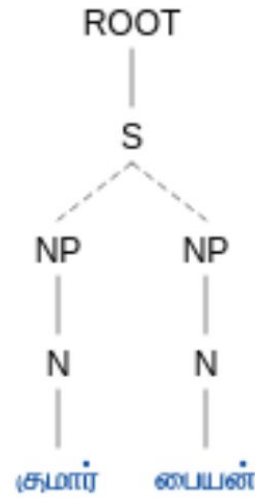
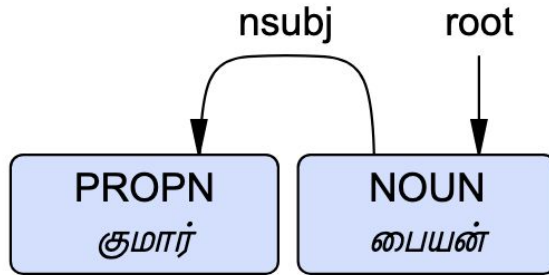
அவன் பந்து
வாங்கினான்
avan pantu vāṅkiṇāṇ
he.NOM ball.**NOM** Bought
He bought **a** ball

அவன் பந்தை வாங்கினான்
avan pantai vāṅkiṇāṇ
he.NOM ball.**ACC** Bought
He bought **the** ball



Null-copula

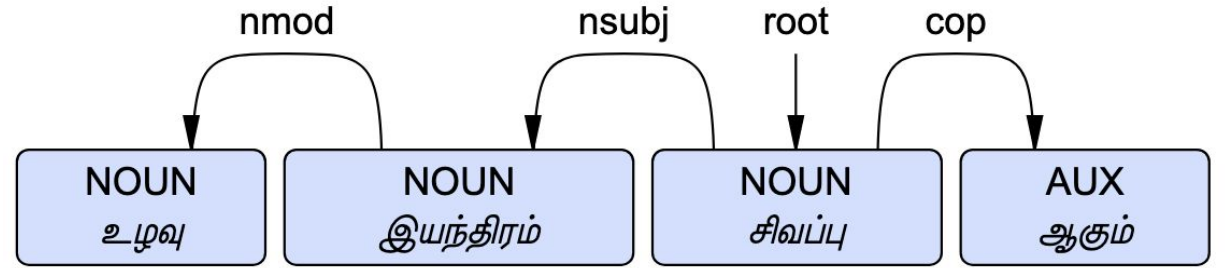
குமார் பையன்
Kumār paiyaṇ
Kumar boy.NOM
'Kumar is a boy.'



| | |
|----------|---|
| PRED | 'ஆகு[5:குமார்], [2:பையன்]>' |
| TNS-ASP | 8 TENSE pres |
| PREDLINK | 2 PRED 'பையன்' RATIONAL +, NUM sg, GEND masc, CASE nom, NTYPE count |
| SUBJ | 5 PRED 'குமார்' NUM sg, GEND masc, CASE nom, NTYPE name |
| 0 | |

Copula

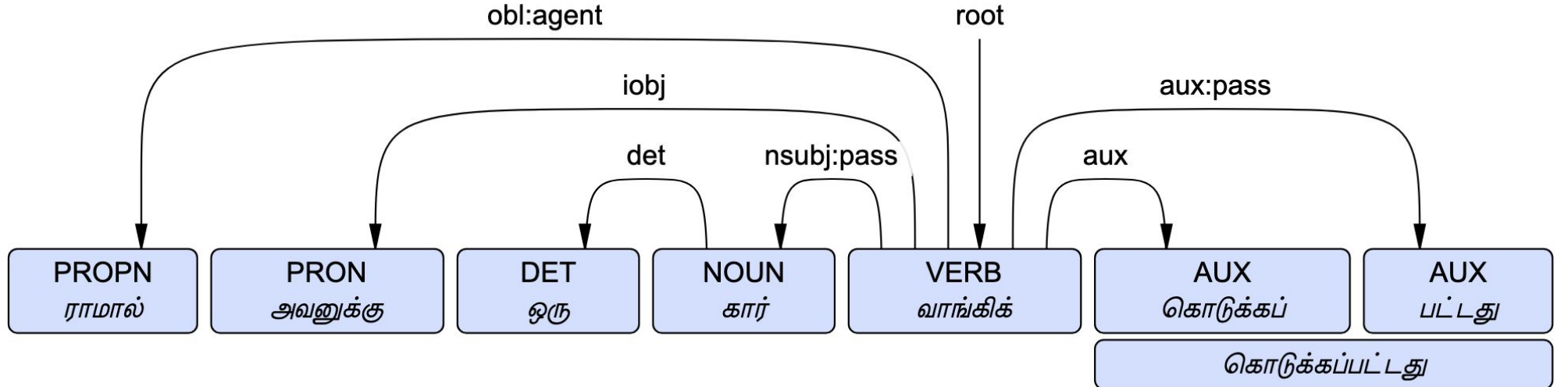
உழவு இயந்திரம் சிவப்பு ஆகும்
uḷavu iyantiram civappu ākum
tillage machine red be
‘The tractor is red.’



Passives - Tamil

ராம் அவனுக்கு ஒரு கார் வாங்கிக் கொடுத்தான்
ram avanukku oru car vangki-k koduththaan
ram.NOM he.DAT a car.NOM buy.VP.SAN give.VP.PAST.3SM
'Ram bought a car for him.'

ராமால் அவனுக்கு ஒரு கார் வாங்கிக் கொடுக்கப்பட்டது
ramaal avanukku oru car vangki-k kodukkappaddathu
ram.INST he.DAT a car.NOM buy.VP.SAN give.INF.SAN.PASS
'A car was bought for him by Ram.'



Causatives - Tamil

- Morphological realisation of Causatives: *vi / pi*

வாங்குவித்தான் (vāṅkuvittān)
வாங்கு -வி -த் -த் -ஆன்
vangu -vi -t -t -aan
buy -CAUS -SAN -PAST -3SMR
'he made somebody buy (something)'

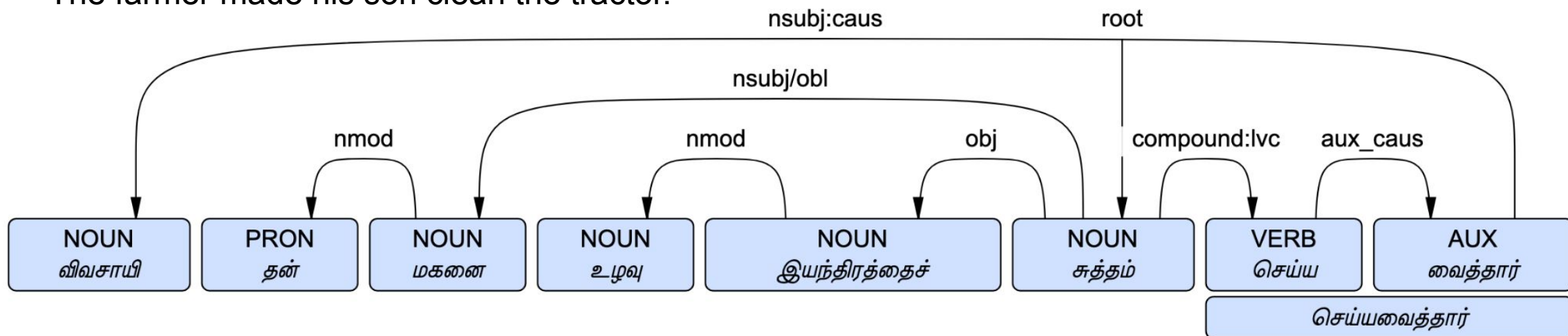
Causatives - Tamil

- Syntactical realisation of causatives: *sei/pannu/vai*

விவசாயி தன் மகனை உழவு இயந்திரத்தைச் சுத்தம்
செய்யவைத்தார்

vivacāyi taṇ makaṇai uḻavu iyantirattaic cuttam ceyyavaittār
farmer.NOM his son.ACC till machine.ACC clean do-make.3SgEpi

The farmer made his son clean the tractor.

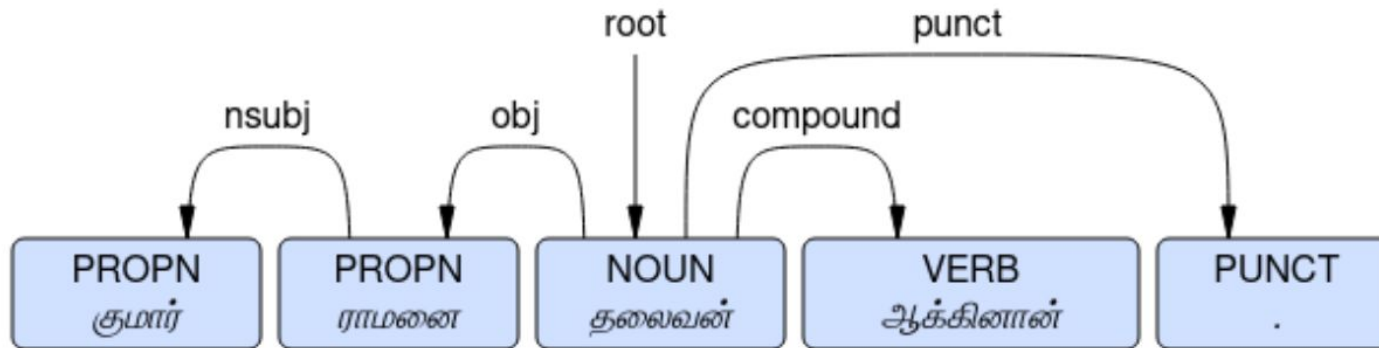


Light verbs (Noun+Verb_{Light}) in Tamil

- Noun + Verb constructions - very productive.

| | | | |
|-----------|----------|------------|----------------|
| குமார் | ராமனைத் | தலைவன் | ஆக்கினான் |
| kumār | rāmanait | talaivan | ākkinān |
| Kumar.NOM | ram.ACC | leader.NOM | make.PAST.3SGM |

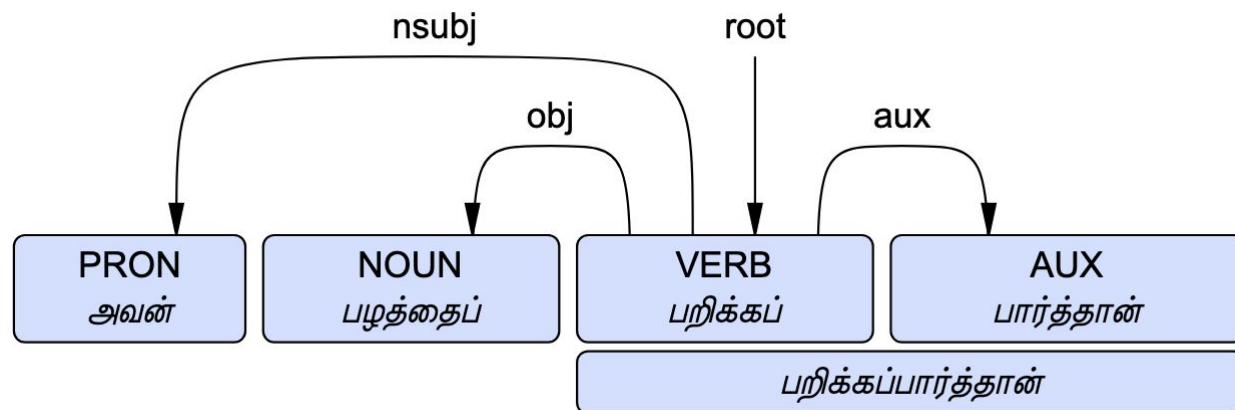
‘Kumar made Raman the leader .’



Light verbs (Verb+Verb_{Light}) in Tamil

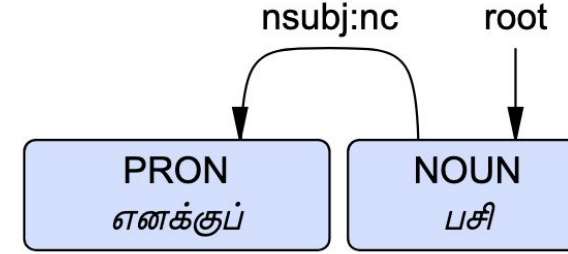
அவன் பழத்தைப் பறிக்கப்பார்த்தான்
avan palattaiṭṭai parikkappārttāṇ
He.NOM Fruit.ACC pick.INF-try.PAST.3SgM
He tried to pick the fruit.

- The verb 'see' express the non-referential meaning of 'try'.
- The light verb as the terminal verbal unit carries functional information, which in this case has to do with tense, number and person.



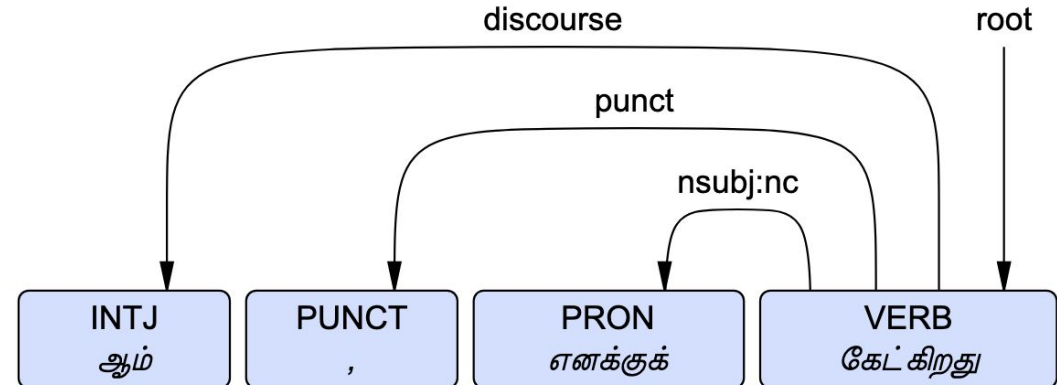
Dative Subjects - Tamil

எனக்குப் பசி
enakkup paci
I.DAT hungry.NOM
'I am hungry.'



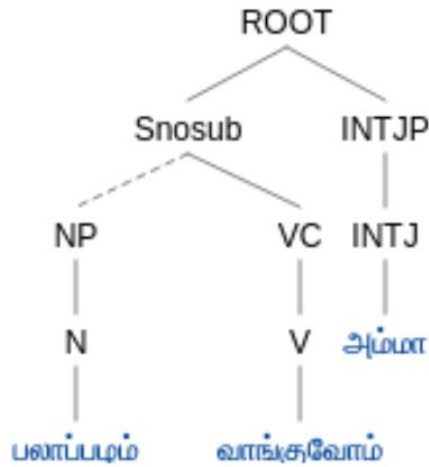
- Agreement between SUB and PRED is always the default agreement - “3SgN”

ஆம், எனக்குக் கேட்கிறது
ām, enakkuk kēṭkiratu
yes, I.DAT hear.PRES.3SN
'Yes, I hear (it)'



Prodrops

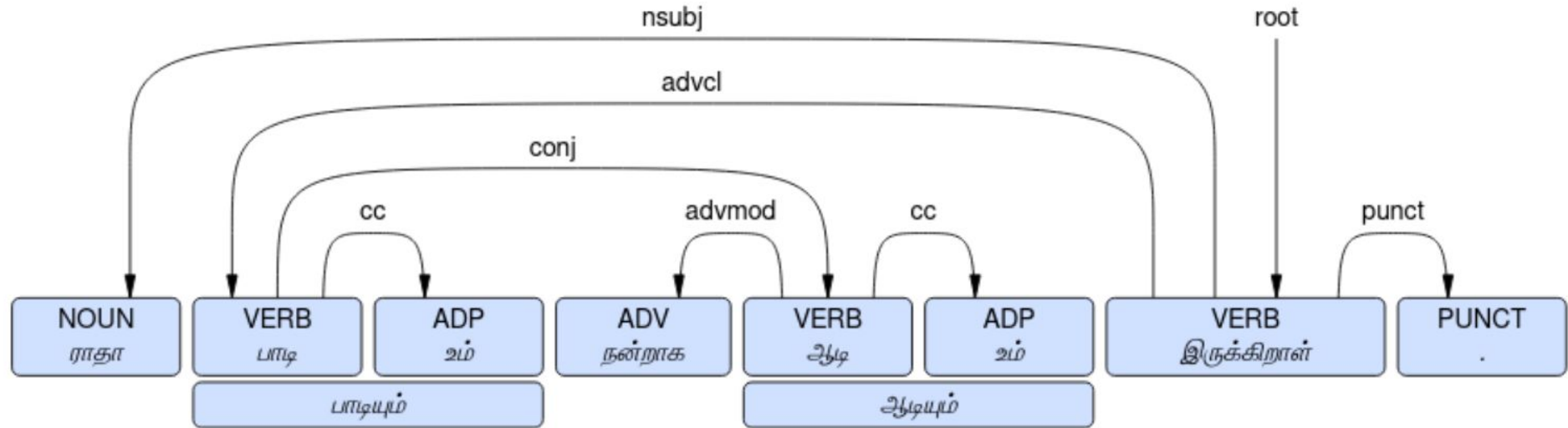
) பலாப்பழம் வாங்குவோம் , அம்மா
 palāppalam vāṅkuvōm am'mā
 Jackfruit buy.FUT.1PLE, mother
 'Mother, we will buy Jackfruit'



| | | | |
|----------------|----------------------------------|--|---------------------|
| PRED | 'வாங்கு<[4:pro], [1:பலாப்பழம்]>' | | |
| TNS-ASP | 5 | TENSE fut, MOOD indicative | |
| OBJ | 2 | PRED | 'பலாப்பழம்' |
| | | NTYPE | NSEM 3 COMMON count |
| | | | NSYN common |
| | | RATIONAL -, PERS 3, NUM sg, HUMAN -, GEND neut, CASE nom, ANIM - | |
| SUBJ | 4 | PRED | 'pro' |
| | | RATIONAL +, PERS 1, NUM pl, GEND epi | |
| VTYP | 0 | main, PASSIVE -, INTERJECTION-LABEL அம்மா | |

Coordination

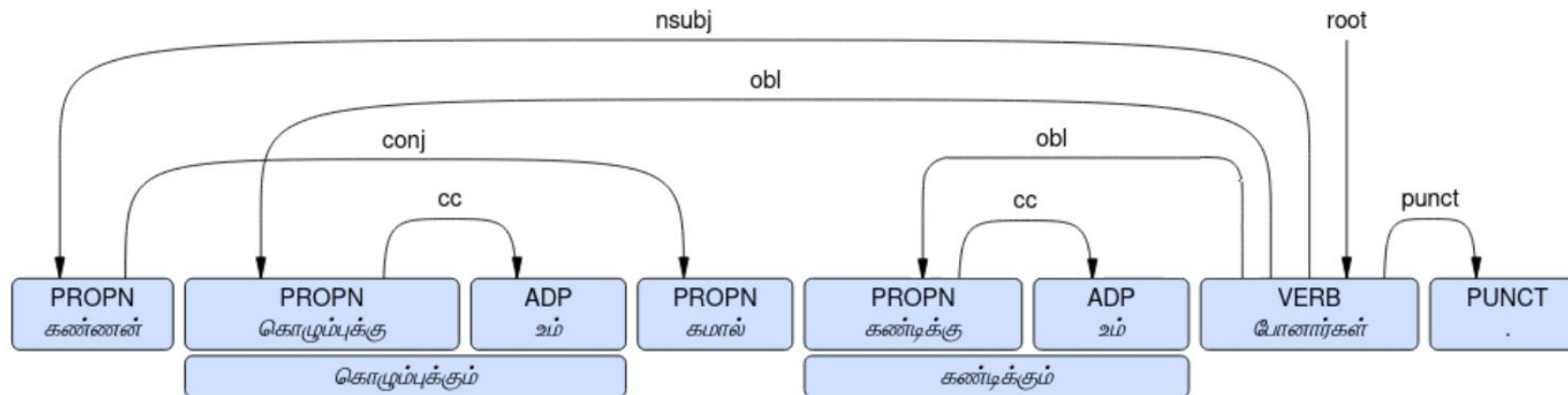
| | | | | |
|-----------------------------------|----------------|----------|-----------------|--------------|
| ராதா | பாடியும் | நன்றாக | ஆடியும் | இருக்கிறாள் |
| rātā | pāṭiyum | nanrāka | āṭiyum | irukkirāl |
| Radha.NOM | sing.VPART-and | well.ADV | dance.VPART-and | be.PRES.2SGF |
| 'Radha has sang and danced well.' | | | | |



Gapping construction

- **A non-projective parse**

| | | | | |
|--|-----------------|-----------|---------------|-------------|
| கண்ணன் | கொழும்புக்கும் | கமால் | கண்டிக்கும் | போனார்கள் |
| kaṇṇan | koḷumpukk-um | kamāl | kaṇṭikk-um | pōnārkaḷ |
| Kannan.NOM | Colombo.DAT-and | Kamal.NOM | Kandy.DAT-and | go.PAST.3PE |
| 'Kannan went to Colombo and Kamal to Kandy.' | | | | |



For more information on Gapping - <https://aclanthology.org/W17-0416.pdf>