A **U**niversal **D**ependencies Treebank for **Gu**jarati

Mayank Jobanputra*, Maitrey Mehta*, Çağrı Çöltekin
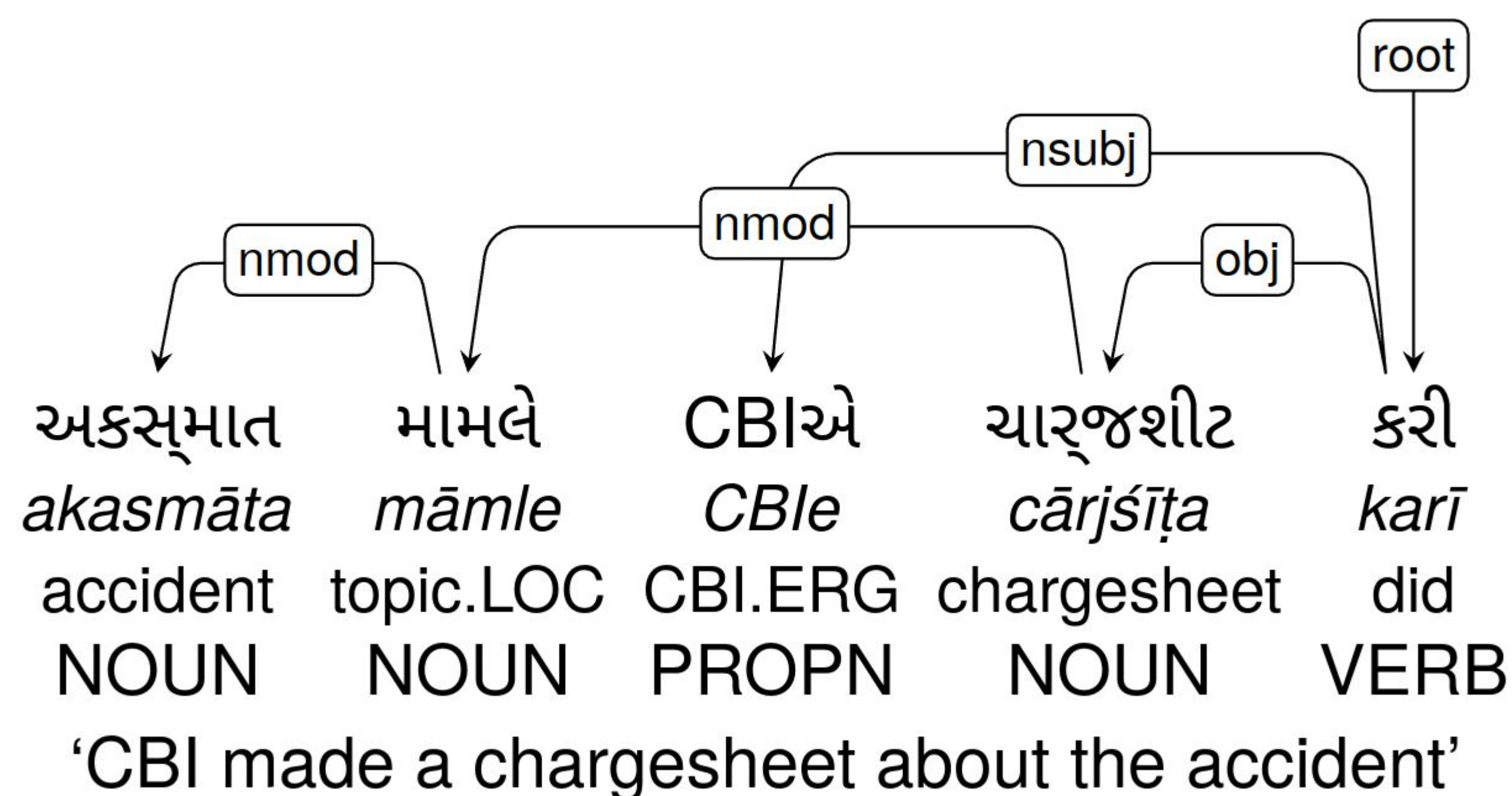
Introduction

- **GujTB**: The first UD treebank in the Indo-Aryan language of Gujarati spoken by over 56M speakers
- 187 sentences with 100 sentences doubly-annotated
- **Diverse Genres** spanning Samanantar (news), UD Cairo (short), and Gujarati translations of the French novella "*Le Petit Prince*" (fiction), and grammar book examples (grammar)

Genre	Sentences	Tokens
news	93	1159
short	20	178
fiction	40	331
grammar	31	217

Data statistics in GujTB

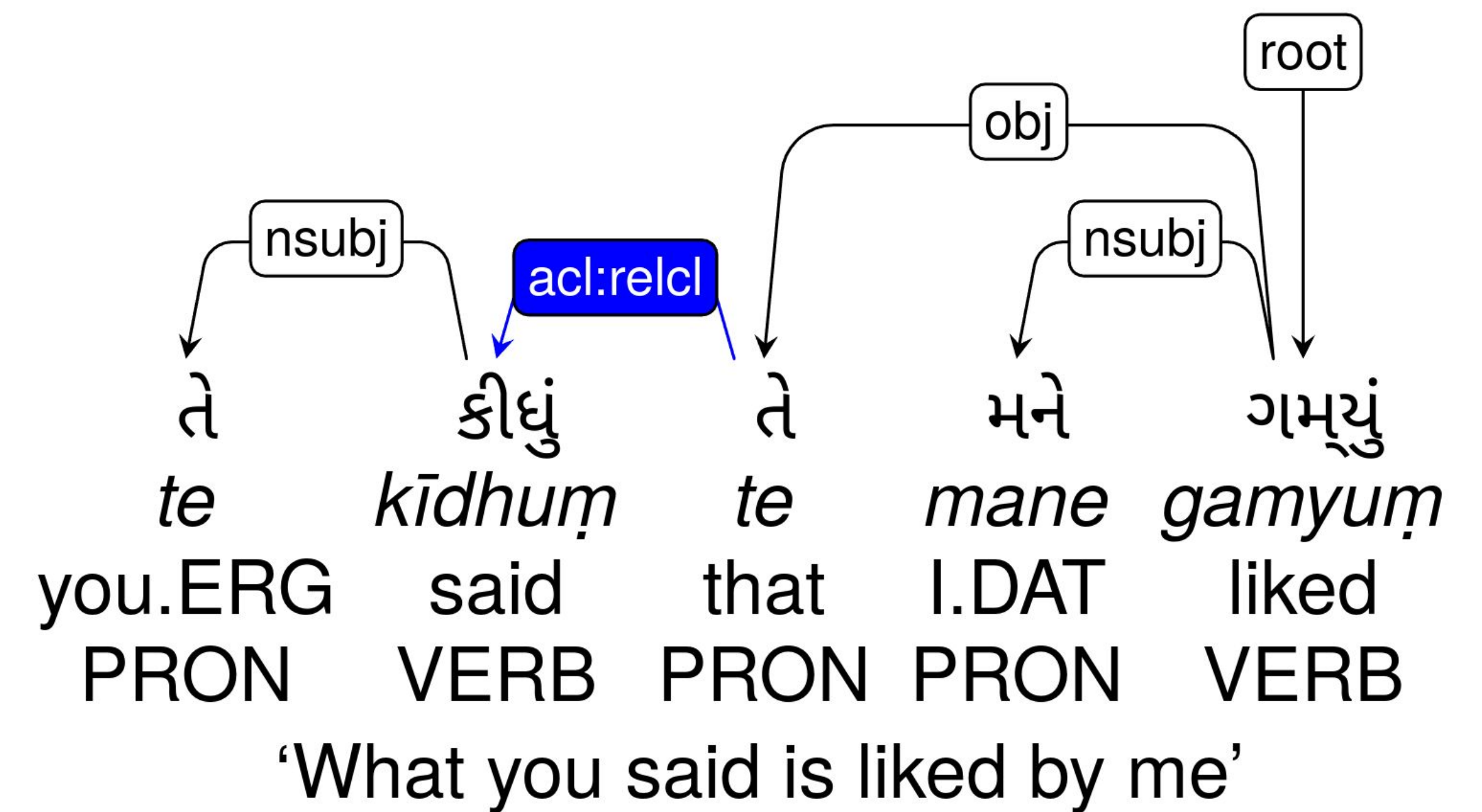- We discuss some interesting cases encountered during the annotation process

Non-projectivty

અકસ્માત	મામલે	CBIએ	ચાર્જશીટ	કરી
akasmāta	*māmle*	*CBIe*	*cārjśīṭa*	*karī*
accident	topic.LOC	CBI.ERG	chargesheet	did
NOUN	NOUN	PROPN	NOUN	VERB

'CBI made a chargesheet about the accident'

- Gujarati allows non-projectivity
- Partial free word order give rise to over-lapping dependency edges

Head-final conjunctions

પીટર	રજત	અને	જેન	સુવર્ણ	જીત્યા
pīṭara	*rajata*	*ane*	*jena*	*suvarṇa*	*jītyā*
Peter	silver	and	Jane	gold	won
PROPN	NOUN	CCONJ	PROPN	NOUN	VERB

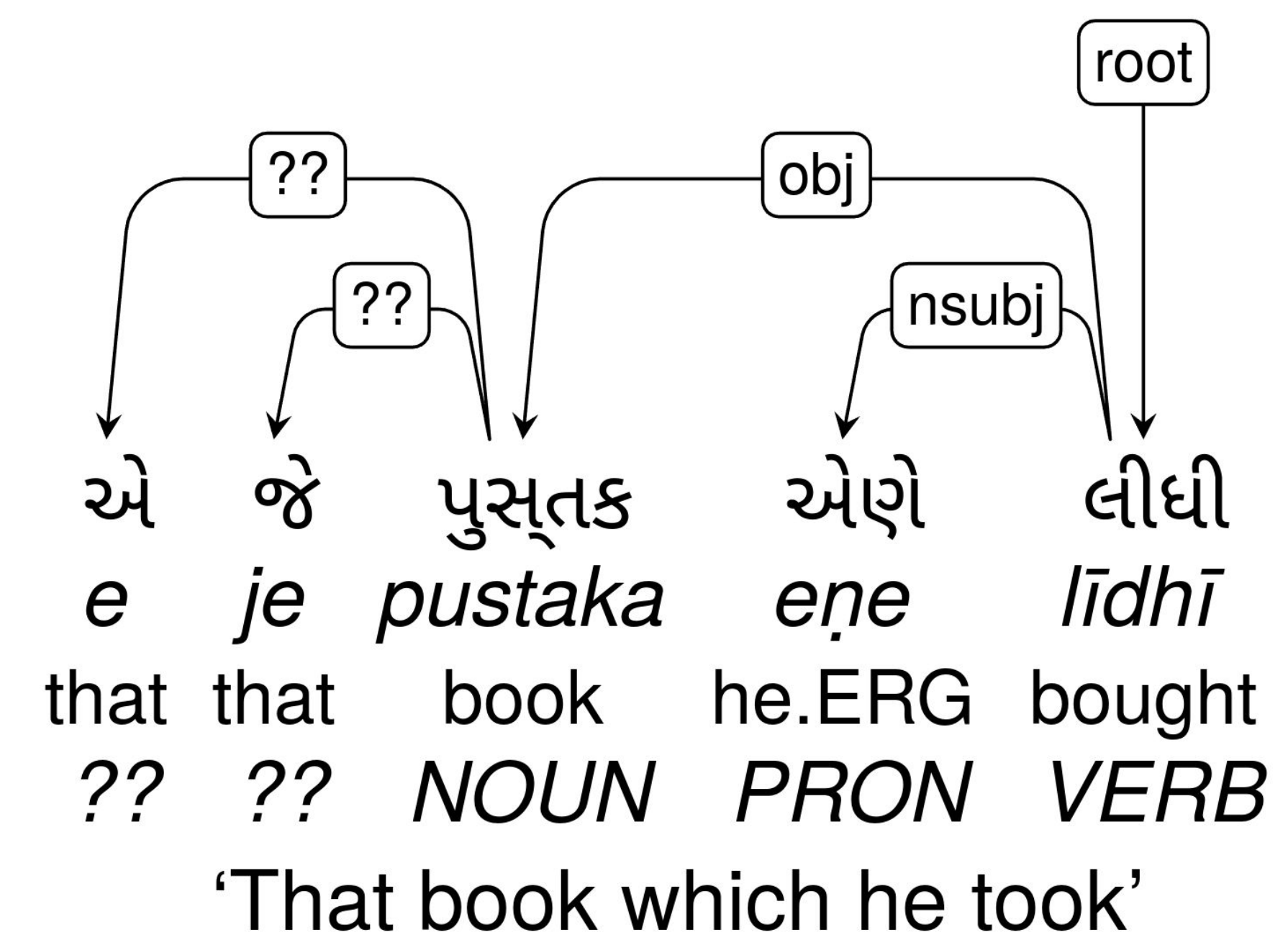'Peter won silver and Jane gold'
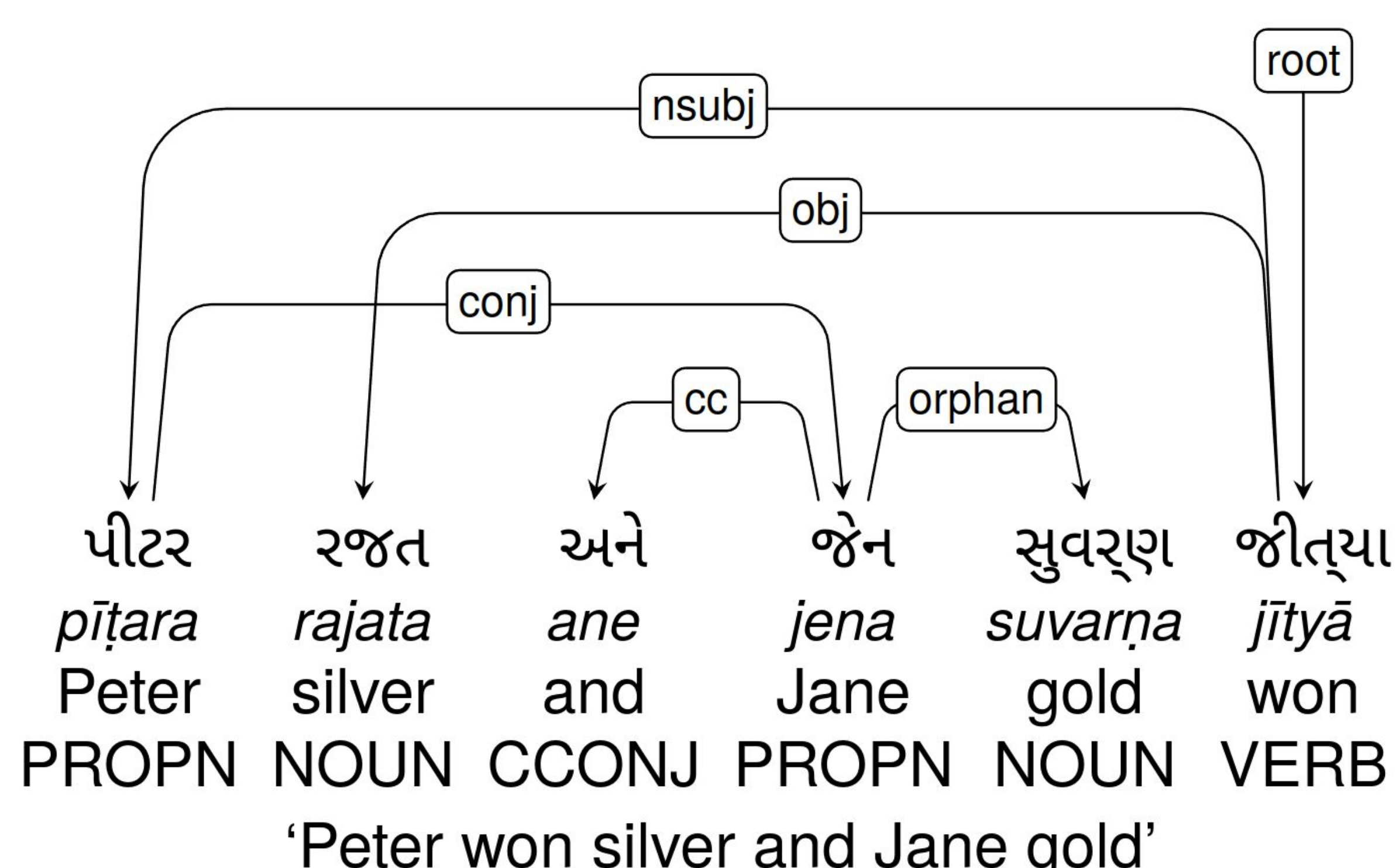
- Gujarati carries case inflections and post-positional attachments on the final conjunct
- '*jītyā*' has a plural inflection requiring it to agree with the head of the coordinated structure

Absence of clausal subjects

તે	કીધું	તે	મને	ગમ્યું
te	*kīdhuṃ*	*te*	*mane*	*gamyuṃ*
you.ERG	said	that	I.DAT	liked
PRON	VERB	PRON	PRON	VERB

'What you said is liked by me'

- A co-referential pronominal '*te*' always present
- A dative nominal construction with experiencer semantics assigned as a '*nsubj*' by prior work

Challenging Construction

એ	જે	પુસ્તક	એણે	લીધી
e	*je*	*pustaka*	*eṇe*	*līdhī*
that	that	book	he.ERG	bought
??	*??*	NOUN	PRON	VERB

'That book which he took'

Multiple Possible Annotations
1. Assigning '*det:predet*' to '*e*' and '*je*' with '*pustaka*' as their head
2. A change in order between '*je*' and '*pustaka*', where '*je*' would act as a subordinating conjunction

The Case for Determiners

- According to Gujarati grammars, demonstrative pronouns like એ, તે, પેલું behave differently when attached to a nominal vs. when used independently
- Gujarati grammars do not discuss determiners as such; yet, the use of nominal-attached pronouns is closer to the UD definition of determiners
- Independently occurring cases treated as pronouns

Links and Contact

Email:
mayank@lst.uni-saarland.de
maitrey@cs.utah.edu
ccoltekin@sfs.uni-tuebingen.de