

# Morpheme-level Coreference Annotations for Pro-dropped Languages



## Motivation and Goal:

- Coreferential relations of dropped pronouns is necessary for Pro-Dropped languages in Coreference Resolution
  - Null-subjects and omitted possessive pronouns
- Information about dropped pronouns are easily deducible from morphology, *morphemes*.
- Representation and Evaluation Scheme

## Approach:

- Each pronominal marker ~ coreferential mention
  - 'Possessive marker' for nouns, and 'Personal marker' for verbs.
- No added any artificially inserted token (e.g. empty node)
  - Eliminates difficulty in determining the most accurate and natural position of the empty node in the sentence
- Validated on Turkish Coreference Resolution

Sen [benim] anne[m]in geldiğ[i]ni gördü[n] mü?

Sen—benim annemin geldiğini gördün mü?

You my mother came see Did

*Did you see the coming of my mother ?*

## Key Observations:

- Representation Scheme: Multiple annotation over a single token is allowed, no need to add artificially inserted token.
- Evaluation Scheme: Pre and post-processors to enhance available CR evaluator to cover dropped pronouns (i.e., multiple annotations over a single word) are developed.