Multilingual Evaluation Campaigns

An update
And a call for participation

Preparatory Meeting Participants

- Agata Savary
- Alina Wróblewska
- Dan Zeman
- Gülşen Eryiğit
- Ivan Vulic
- Joakim Nivre

- Leonie Weissweiler
- Louis Esteve
- Omer Goldman
- Reut Tsarfaty
- Tuğba Pamay
- Tanja Samardzic

Evaluation Campaign in Focus

- Morpho-Syntactic Analyso-Parsing
  - Presented by Omer, Leonie and Reut

- White-space independent morpho-syntactic trees

- Natural annotation for polysynthetic languages

- Practically: adding phrase-level features for UD trees
  - Morpho-syntactic trees to be accessible by dropping all nodes with no phrase-level feats

For those who follow social media transitions on Capitol Hill, this will be a little different.

UD tree

MSAP tree
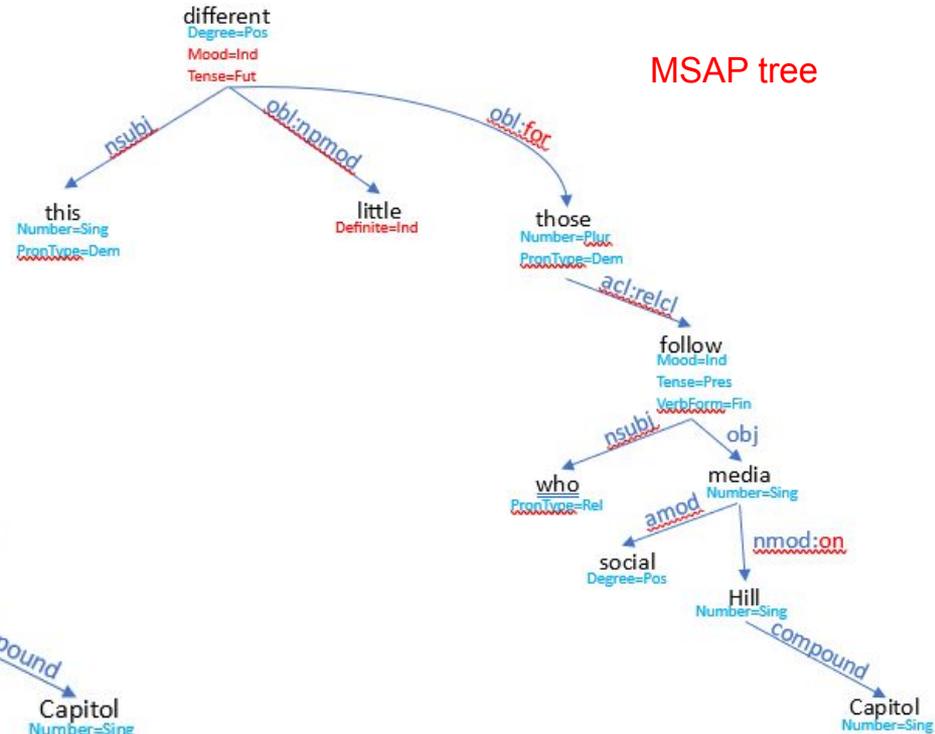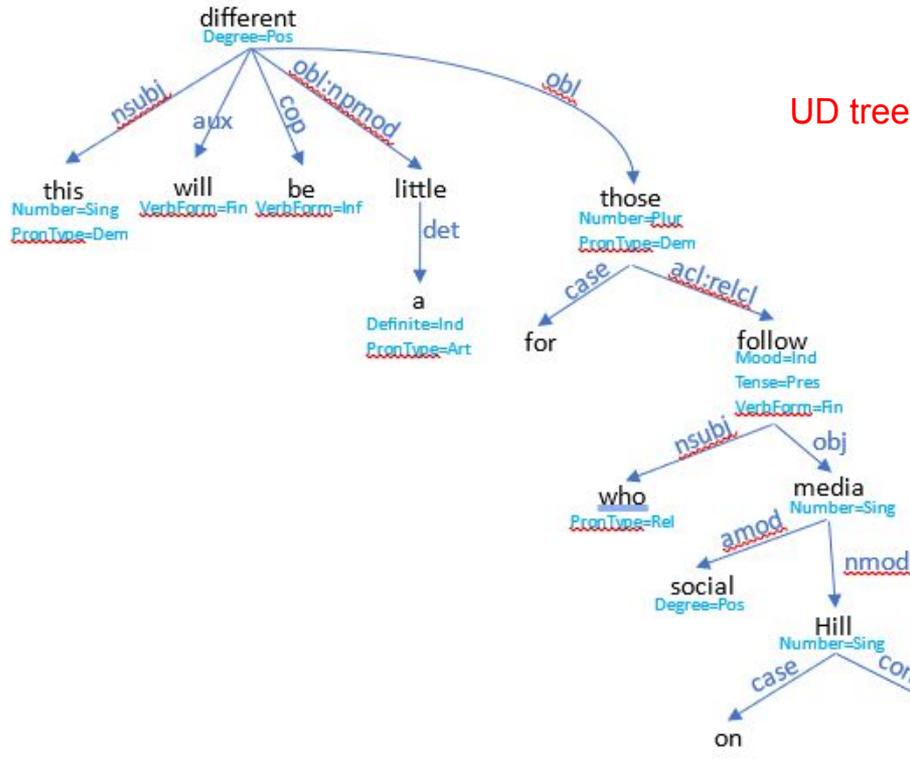
Next Steps

- Finalizing the schema

- Collecting data
    - Converting UD treebanks
    - Data for new languages

Next Steps

- Finalizing the schema
  - Forming a discussion group
  - Raising and solving potential issues
  - As multilingual as possible
  - End date: February's Naples meeting

- Collecting data
  - Converting UD treebanks
  - Data for new languages

Next Steps

- Finalizing the schema
  - Forming a discussion group
  - Raising and solving potential issues
  - As multilingual as possible
  - End date: February's Naples meeting

- Collecting data
  - Converting UD treebanks
  - Data for new languages
  - Official annotation guidelines
  - Per-language annotation teams
  - Central validation process
  - Shared-task: beginning of 2025 at most

Next Steps

- Finalizing the schema
  - Forming a discussion group ← today!
  - Raising and solving potential issues
  - As multilingual as possible
  - End date: February's Naples meeting

- Collecting data
  - Converting UD treebanks
  - Data for new languages
  - Official annotation guidelines
  - Per-language annotation teams
  - Central validation process
  - Shared-task: beginning of 2025 at most

Discussion Group

- Goal: redefine dependency trees to be white-space independent
- Means: a new column in the CoNLL-U file - phrase-level features
  - Defined only for content lexemes
  - Based on morphological features and function lexemes

Discussion Group

- Goal: redefine dependency trees to be white-space independent
- Means: a new column in the CoNLL-U file - phrase-level features
  - Defined only for content lexemes
  - Based on morphological features and function lexemes

- Problem: these bullets are not sufficient to define the data structure

Discussion Group

- Goal: redefine dependency trees to be white-space independent
- Means: a new column in the CoNLL-U file - phrase-level features
  - Defined only for content lexemes
  - Based on morphological features and function lexemes

- Problem: these bullets are not sufficient to define the data structure
- Example: ADPs and case morphemes should be treated the same
  - Is it **ADP → case feature** or **cases → ADP-like node**?
  - What about coverbs?

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use
- Distribution of a "worksheet"
  - List of sentences annotated manually by all participants

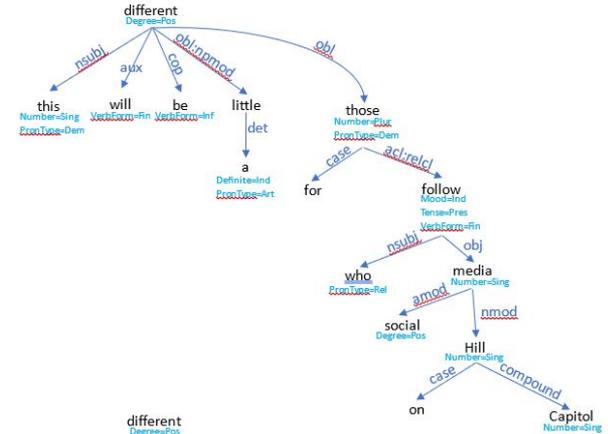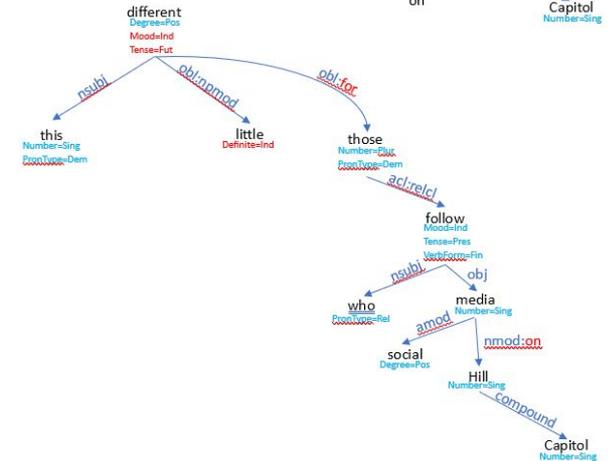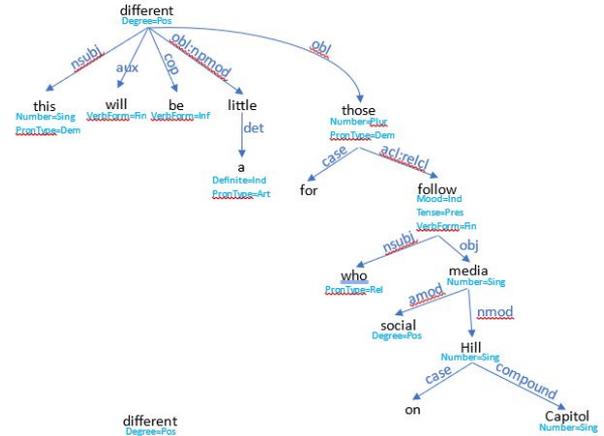For those who follow social media transitions on Capitol Hill, this will be a little different. **sentence**

UD tree

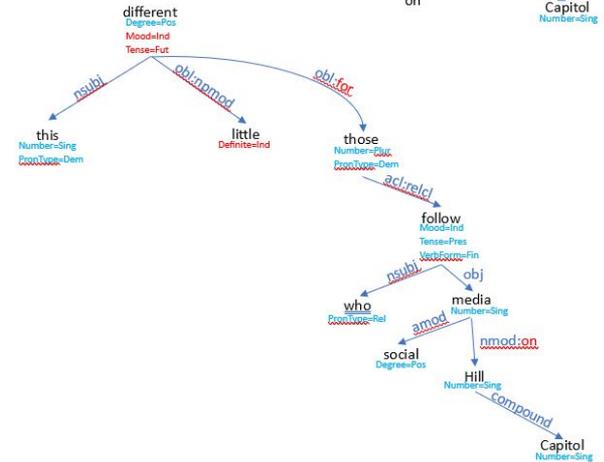**MSAP tree**

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use
- Distribution of a "worksheet"
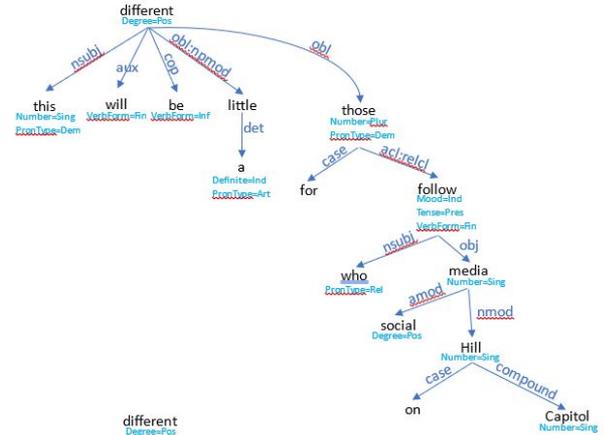 - List of sentences annotated manually by all participants

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use
- Distribution of a "worksheet"
  - List of sentences annotated manually by all participants
- Discussion based on annotations
- Documentation of secondary issues
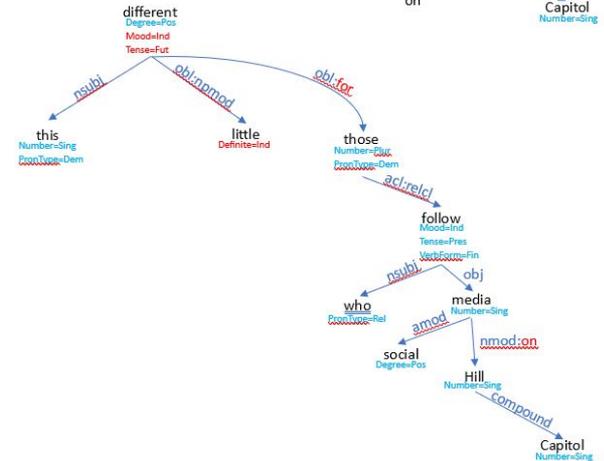- Finalizing the schema

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use
- Distribution of a "worksheet"
  - List of sentences annotated manually by all participants
- Discussion based on annotations
- Documentation of secondary issues
- Finalizing the schema

- Most discussions offline
 - Hopefully there will be many participants

Modus Operandi

- Collection of issues to be treated
  - All definitions-by-whitespace in current use
- Distribution of a "worksheet"
  - List of sentences annotated manually by all participants
- Discussion based on annotations
- Documentation of secondary issues
- Finalizing the schema

- Most discussions offline
  - Hopefully there will be many participants
- Online "workshop" for annotating together

sentence

UD tree
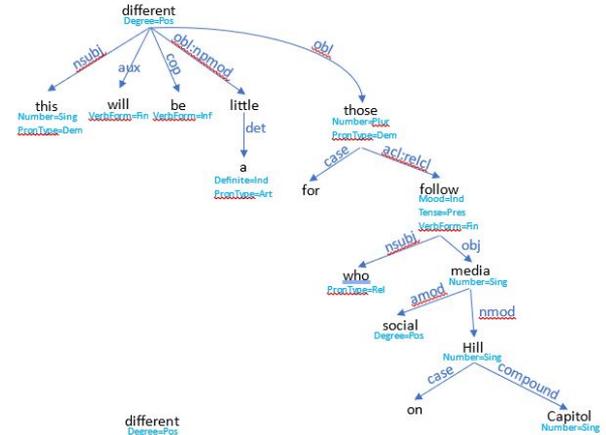
MSAP tree
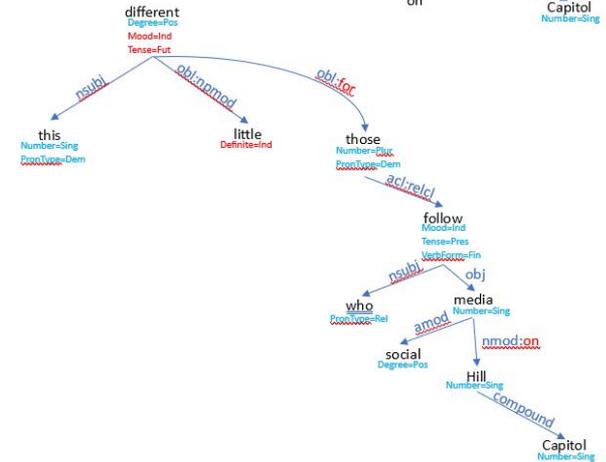
Call for Action and Participation

- Want to take part?
  - Email: omer.goldman@gmail.com
- Know someone in- or outside UniDive that may be interested?
  - Email: omer.goldman@gmail.com
- Starting date: very soon
  - Schema should be final by February

WE WANT YOU!