

Standardizing Annotations in Turkic Languages

Objective: Standardize annotations in Universal Dependencies treebanks of Turkic languages.

Introduction:

- Growing number of treebanks in UD.
- Importance of annotation consistency.

UD Turkic Workshop 2023 (UDTW23):

- September 2023, Istanbul.
- Discussions on issues and examples.

Issues Discussed:

- Tokenization, tense/mood tags, oblique/object distinction.
- Question particle, code switching, transcription, pronominalized nouns.

Conclusions:

- Workshop as a catalyst for unified annotation approach.
- Future steps: comprehensive paper on issues and decisions.

References and Acknowledgements: See poster for details.

Languages in UD: Kazakh, Kyrgyz, Tatar, Turkish (9), Uyghur, Yakut, Old Turkish, and Turkish-German

Coming to UD: Uzbek, Ottoman Turkish, Turkish (1)