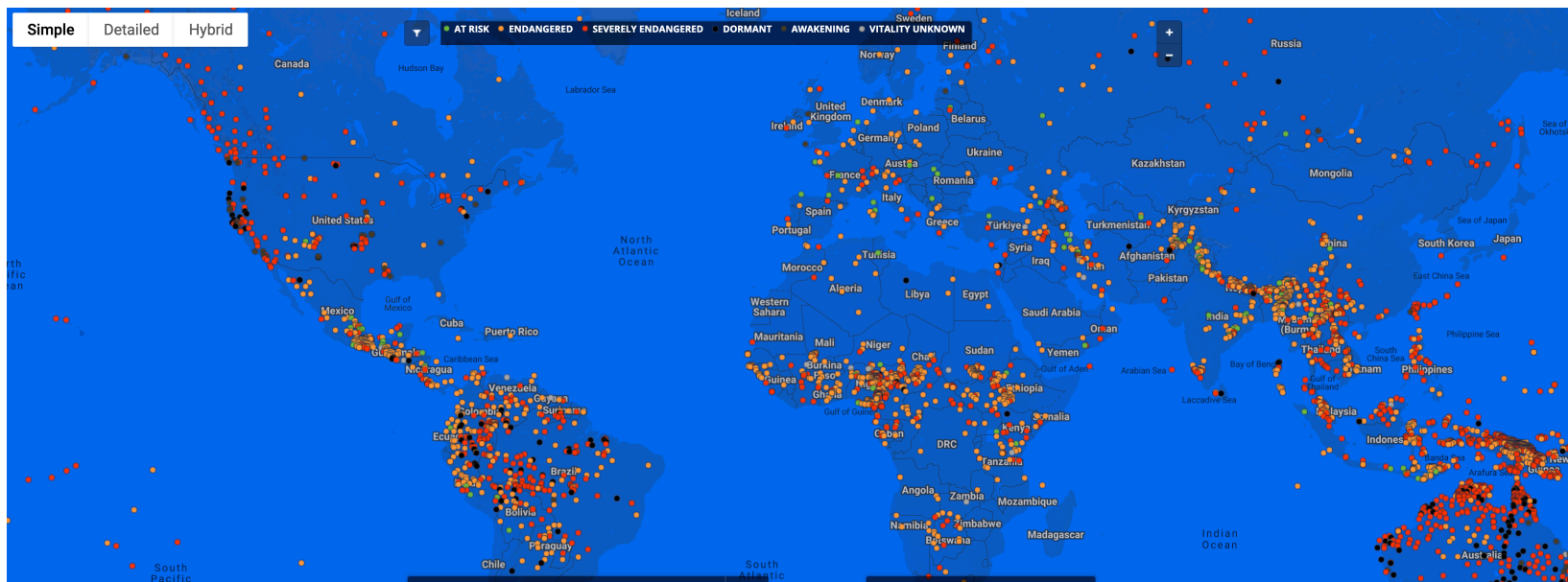


Natural Language Processing

Info 159/259

Lecture 20: NLP for Low-Resource Languages

Languages of the World

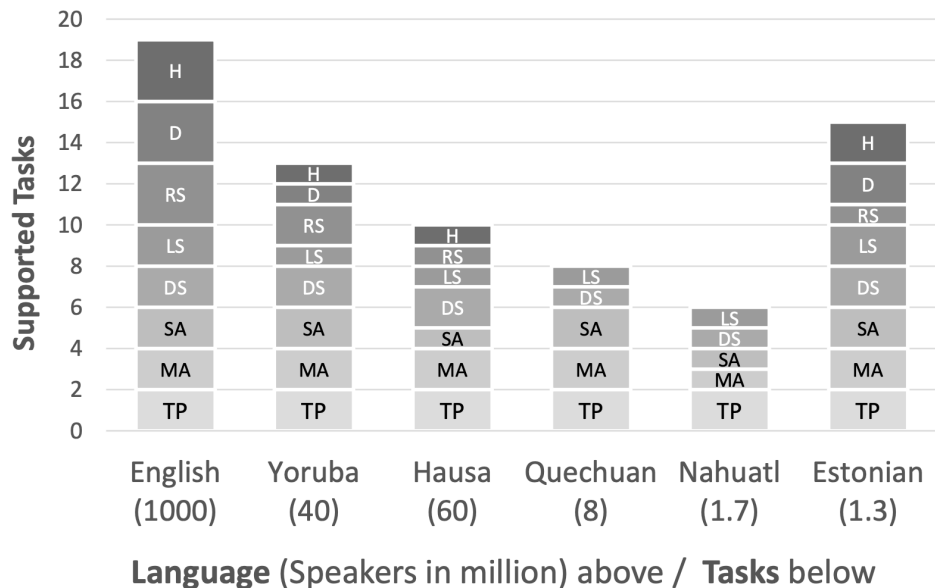


<https://endangeredlanguages.com/>

NLP Beyond English

- Cultural differences get reflected in language diff.
- Linguistic Diversity present in different layers

NLP Beyond English



- H: Higher-level NLP applications
- D: Discourse
- RS: Relational semantics
- LS: Lexical semantics
- DS: Distributional semantics
- SA: Syntactic analysis
- MA: Morphological analysis
- TP: Text processing

Challenges: Word Boundaries

- CJK languages
-

Challenges: Morphology

- Arabic case study
-

Challenges: Syntax

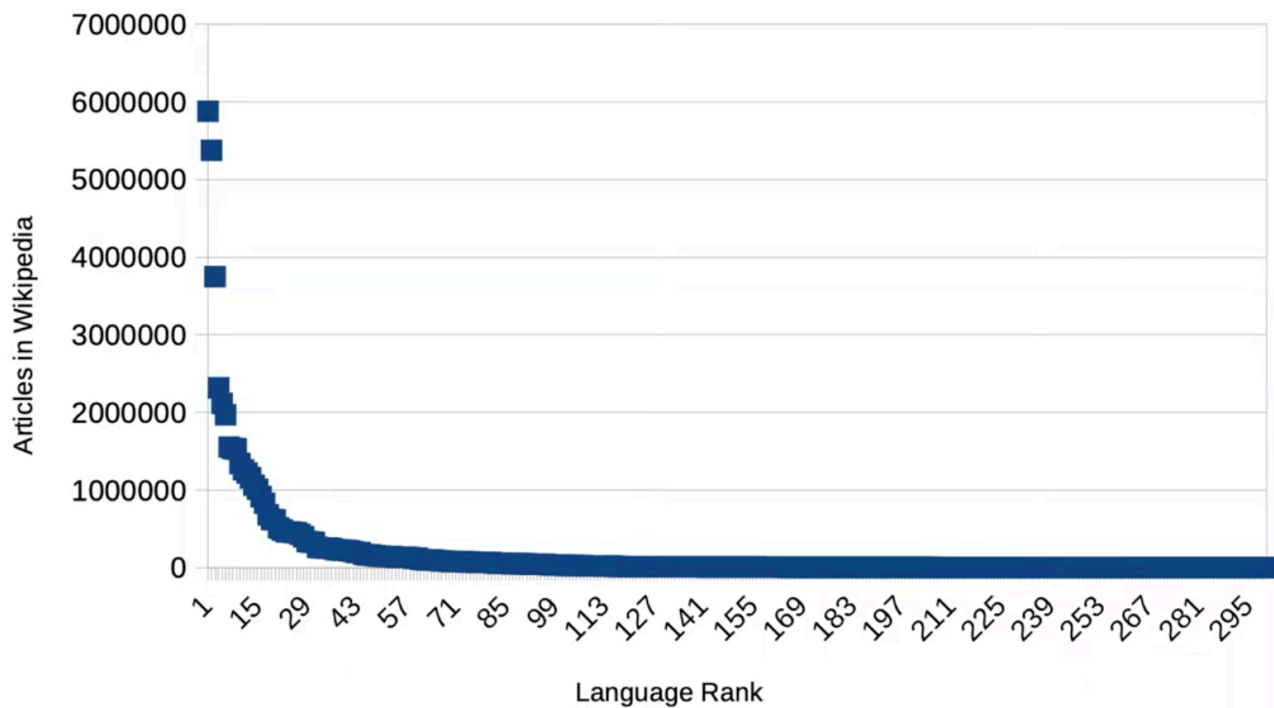
Challenges: NER & Transliteration

- Transliteration of names across languages
-

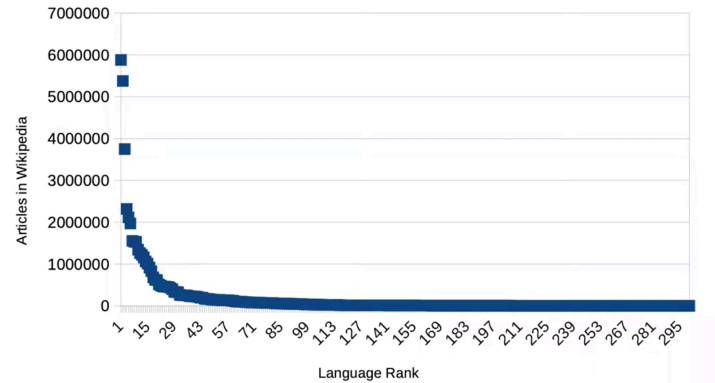
Aspects of “Low Resource”

How Low is *Low Resource*

Data



Data



Generating Additional Data

Data Augmentation

Distant & Weak Supervision

Cross Lingual Projection

- M

Noise In the Data

- M

Transfer Learning

Low Resource MT

Low Resource MT

Multilingual Models

Multilingual Embeddings

Multilingual LMs

Low Resource Languages on LLMs

Low Resource Languages on LLMs