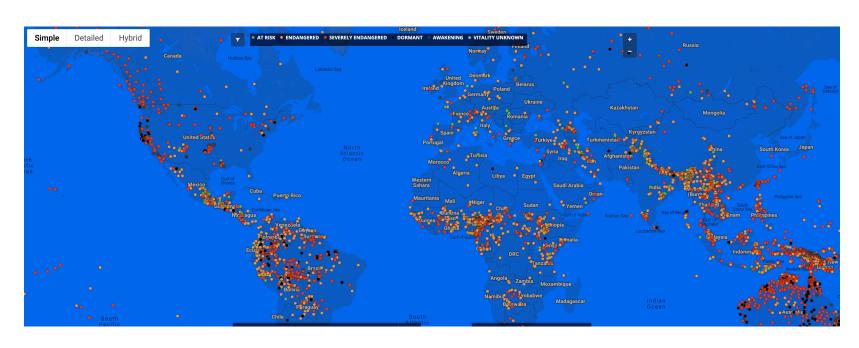
#### Natural Language Processing

Lecture 22: NLP for Low-Resource Languages

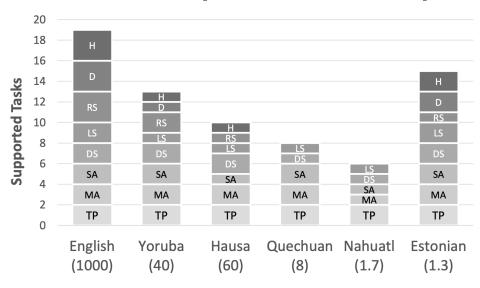
# Languages of the World



### NLP Beyond English

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

#### NLP Beyond English



**Language** (Speakers in million) above / **Tasks** below

- H: Higher-level NLP applications D: Discourse
- RS: Relational semantics
- DS: Distributional semantics
- MA: Morphological analysis

- LS: Lexical semantics
- SA: Syntactic analysis
- TP: Text processing

#### Challenges: Word Boundaries

CJK languages

•

## Challenges: Morphology

Arabic case study

•

## Challenges: Syntax

#### Challenges: NER & Transliteration

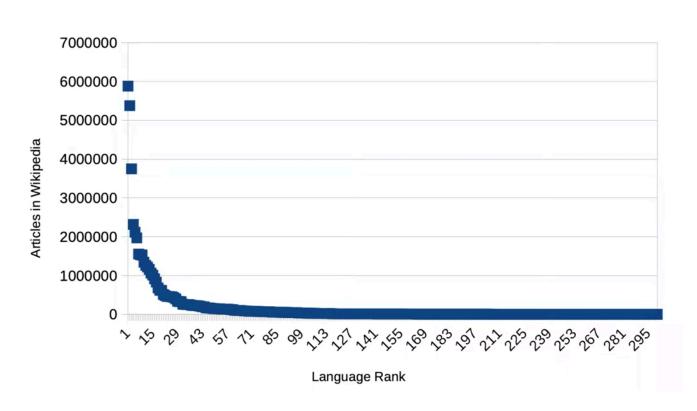
Transliteration of names across languages

•

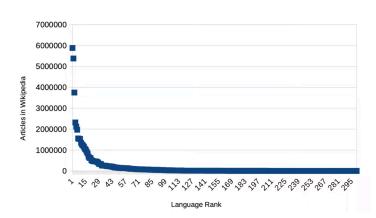
#### Aspects of "Low Resource"

#### 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