Dependency Parsing

Nathan Green

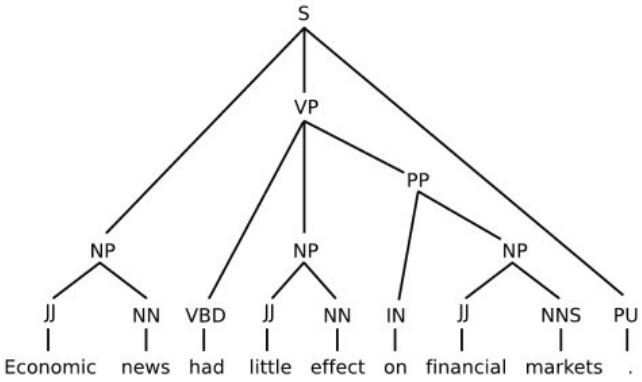
MFF Charles University

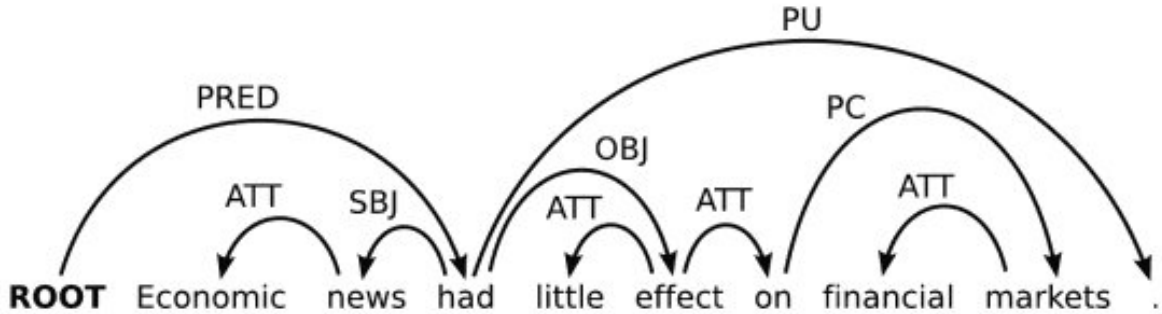
June 2, 2011

Outline of my talk

- Motivation
- Types of Dependency Parsers
- Research Directions
- Conclusion

Why dependency structures?





- Constituent trees have some short comings, model mainly phrasal adjacency
- Useful in sentence analysis, can be represent agreement more appropriately
- Preferred in free word order languages such as Czech

What is a head and what is a dependent?

- The head determines the syntactic category of a construction, and can sometimes replace the construction
- The head determines the semantic category of a construction, and the dependent gives the semantic specification
- The head is obligatory, the dependent is optional
- The head selects the dependent and determines whether the dependent is needed
- The form of the dependent depends on the head (agreement)
- The linear position of the dependent is specified with reference to the head

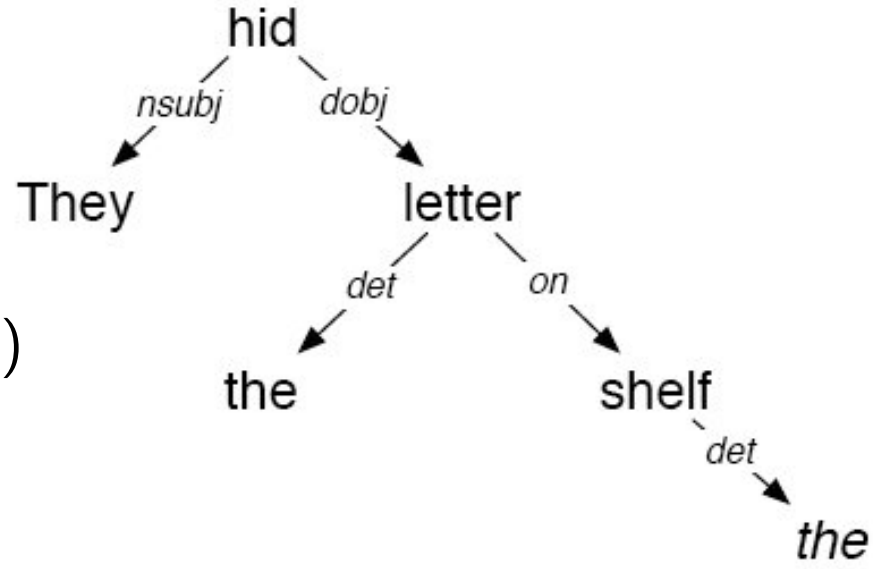
Different annotation standards

The task is essentially ill-defined and left to heuristics in many cases. Different annotation schema require different constructions of the dependency relation.

Much of the current advancement started at shared task competitions

- CoNLL Shared task dedicated to multi-lingual dependency parsing in 2006
- Later years also added domain adaptation
- 13 Languages from 7 language families
- Scores ranged from the 60% to the 90% accuracy depending on language and data sizes

How do we measure success



- *U*nlabeled *A*ttachment *S*core (UAS)
- *L*abeled *A*ttachment *S*core (LAS)
- Compete Sentence Accuracy

3 major approaches to Dependency Parsing

- Transition Based
  - Malt Parser
  - Z-par
- Graph Base
  - MSTParser
- Constituent Conversion
  - Pennconverter
  - Stanfordconverter

Transition Based Parsing

- Pros
  - Fast and efficient algorithms
- Cons
  - Usually Greedy/ No global optimal search
  - Error propagation is a problem especially in searches with long arc paths

Graph Based Parsing

- Pros
  - Global optimal search
  - Tends to perform well on longer sentences since it's search is exact
- Cons
  - Poor feature representations
  - Increasing parameters hurts parsing efficiency

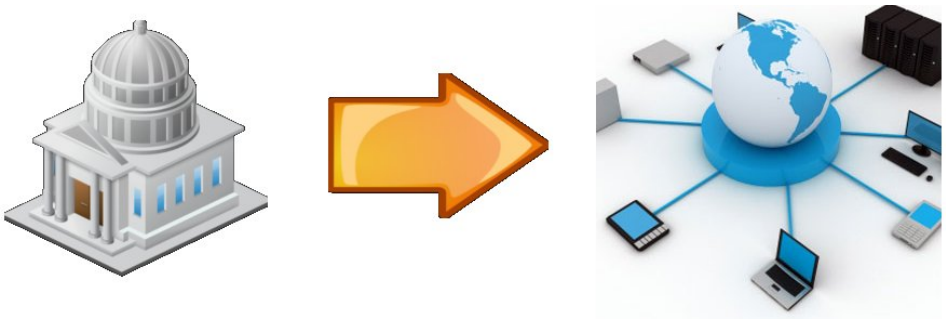
Constituent Based Parsing

- Use state-of-the-art Constituent Parsers
- "Transform" the parse into a dependency parse
- Returns state-of-the-art dependency parsing but only tried in English on corpora that weren't dependency based to begin with.

What is the goal of new dependency parsing research

Improve Parsing Accuracy through domain modification and annotation structure

Domain Adaptation

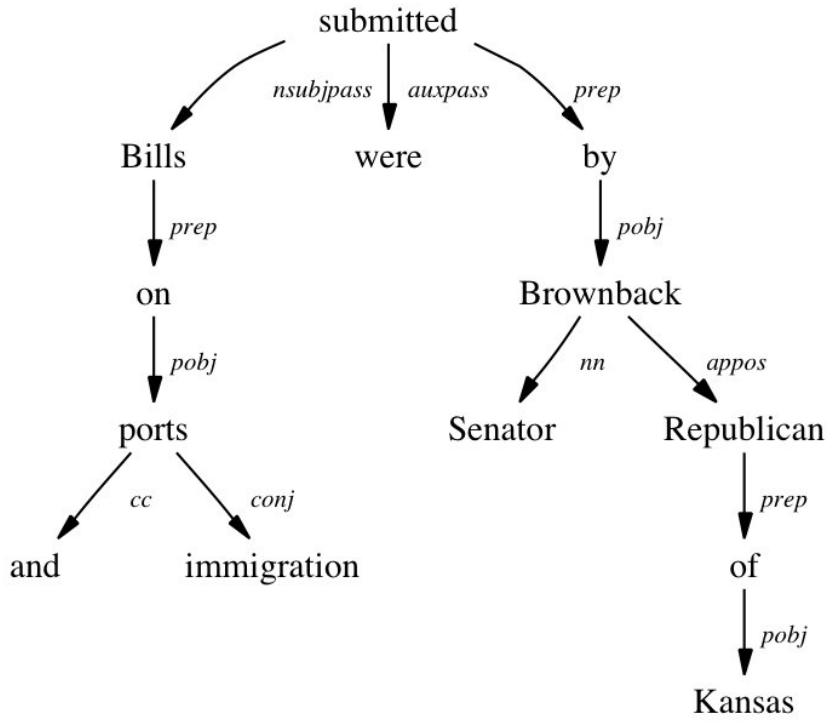


Domain adaptation is the task of taking a parsing model training on one domain and applying it to an unrelated domain. This usually has both grammar and lexicon implications.

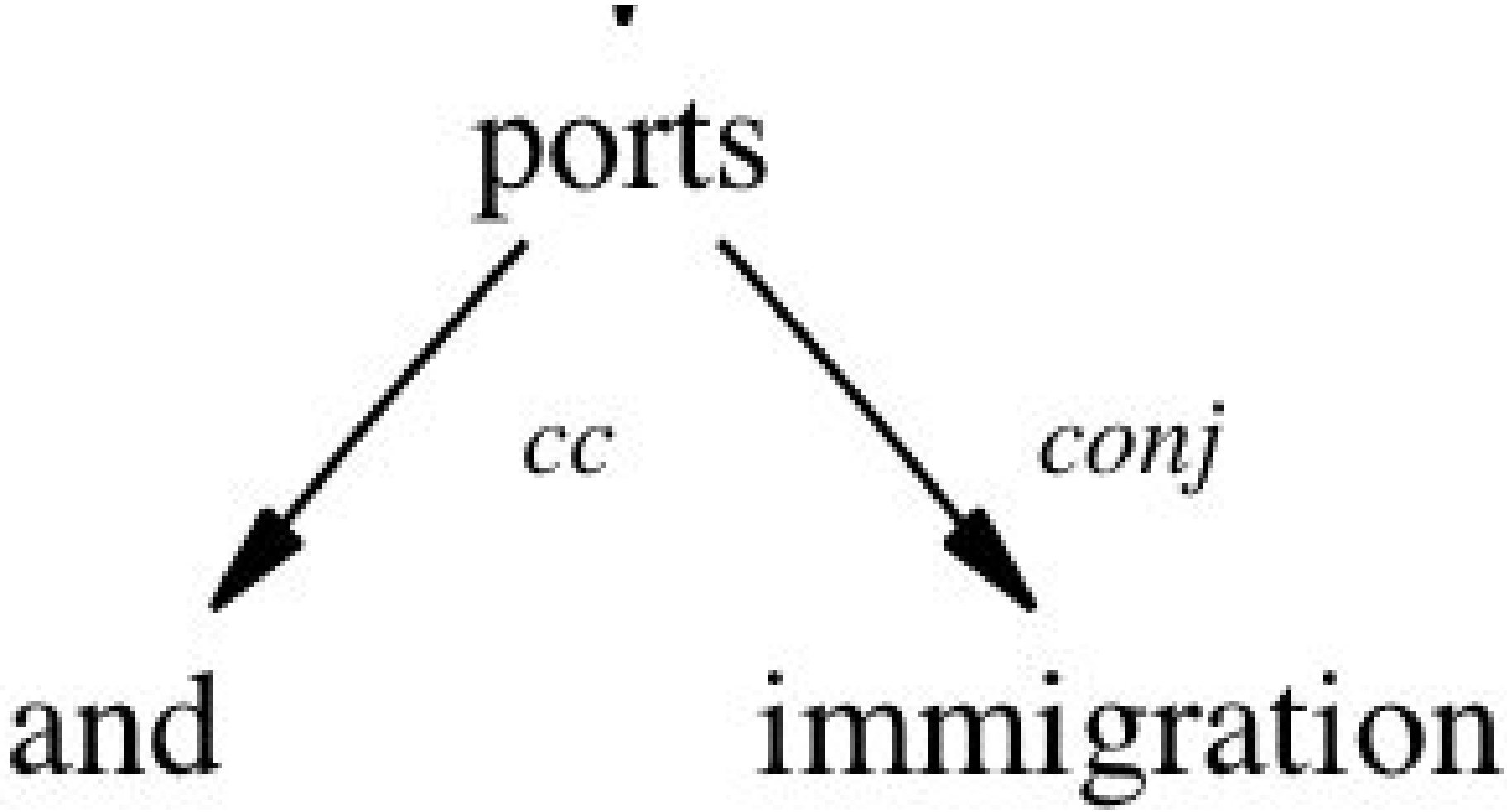
- Domain adaptation has been shown to be effective with:
  - Up-training
  - Self training
  - Model selection

Annotation Structure

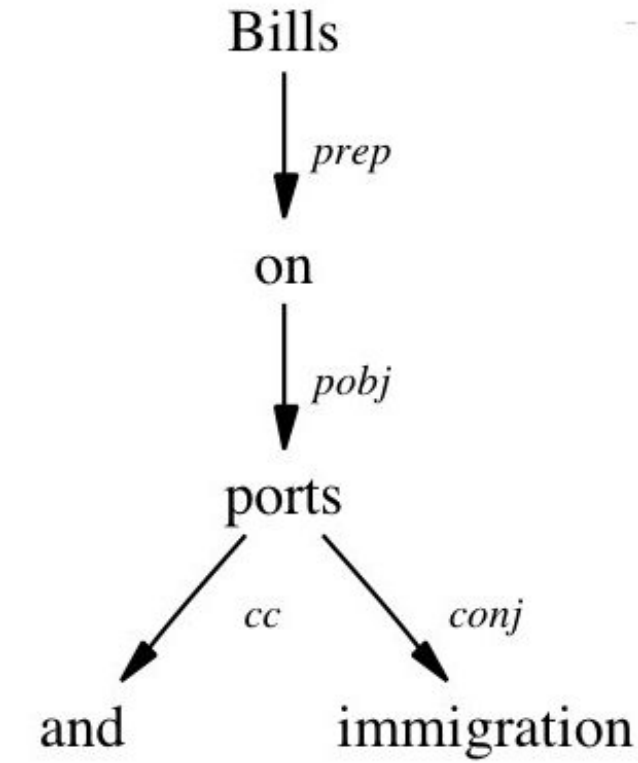
Coordination and head selection can play an important role in parsing



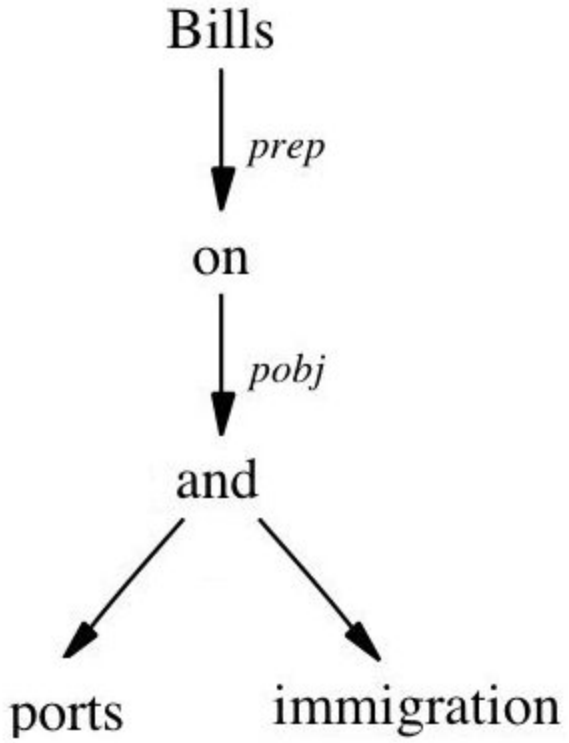
Annotation Structure



Coordination along with other annotations may be handled very differently

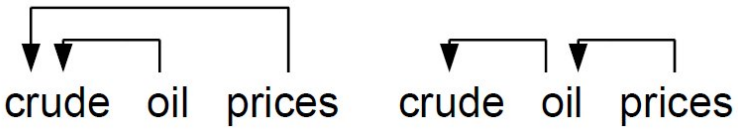


Stanford Coordination



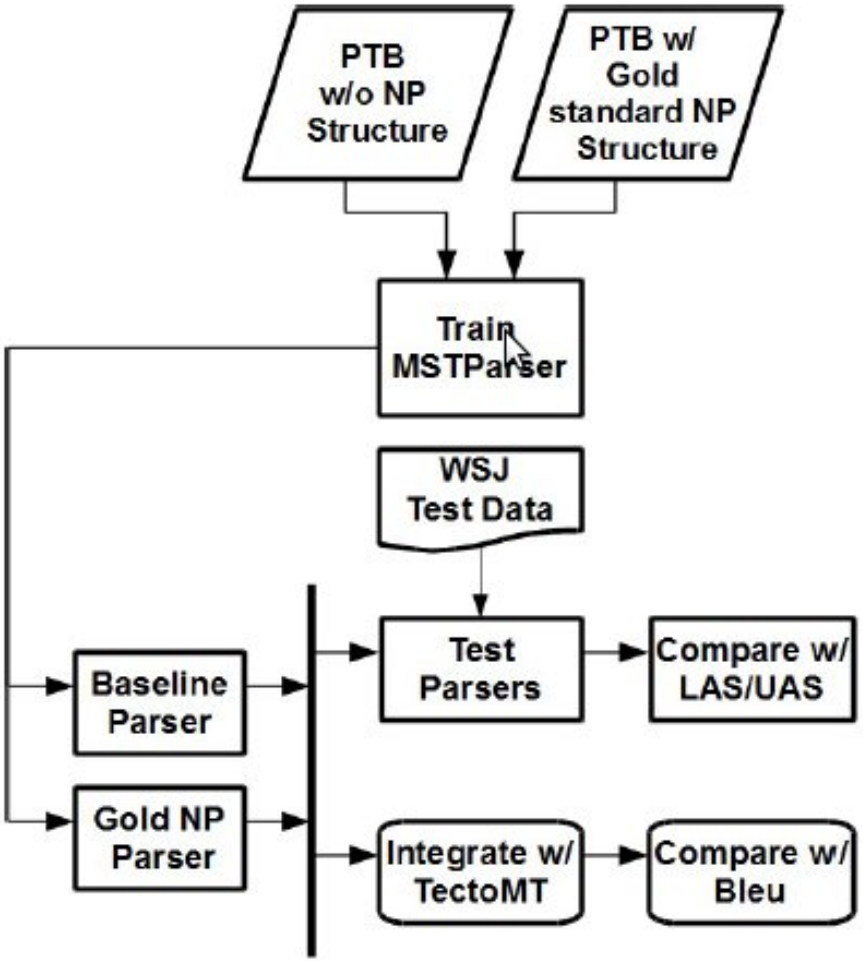
PDT Coordination

Noun Phrase annotation guidelines impact other applications in the NLP Pipeline



- Noun Phrases contain possible ambiguity.
- Most current approaches treat Noun Phrase structure as flat
- Gold standard Noun Phrase annotation exist from Vadas and Curran
- Does this additional structure help either dependency parsers or NLP applications that make use of dependency structures?

Noun Phrase annotation guidelines impact other applications in the NLP Pipeline



Noun Phrase results on Machine Translation (TectoMT)

Systems	Bleu
Baseline Parser	9.47
Gold NP Parser	**9.70**

- NP structure had little effect on UAS and LAS
- Bleu score did improve with statistical significance using pairwise evaluation (95% confidence value)
- Shows that maybe Parsing should be evaluated in the NLP pipeline and not just with accuracy measures.

Where to go from here

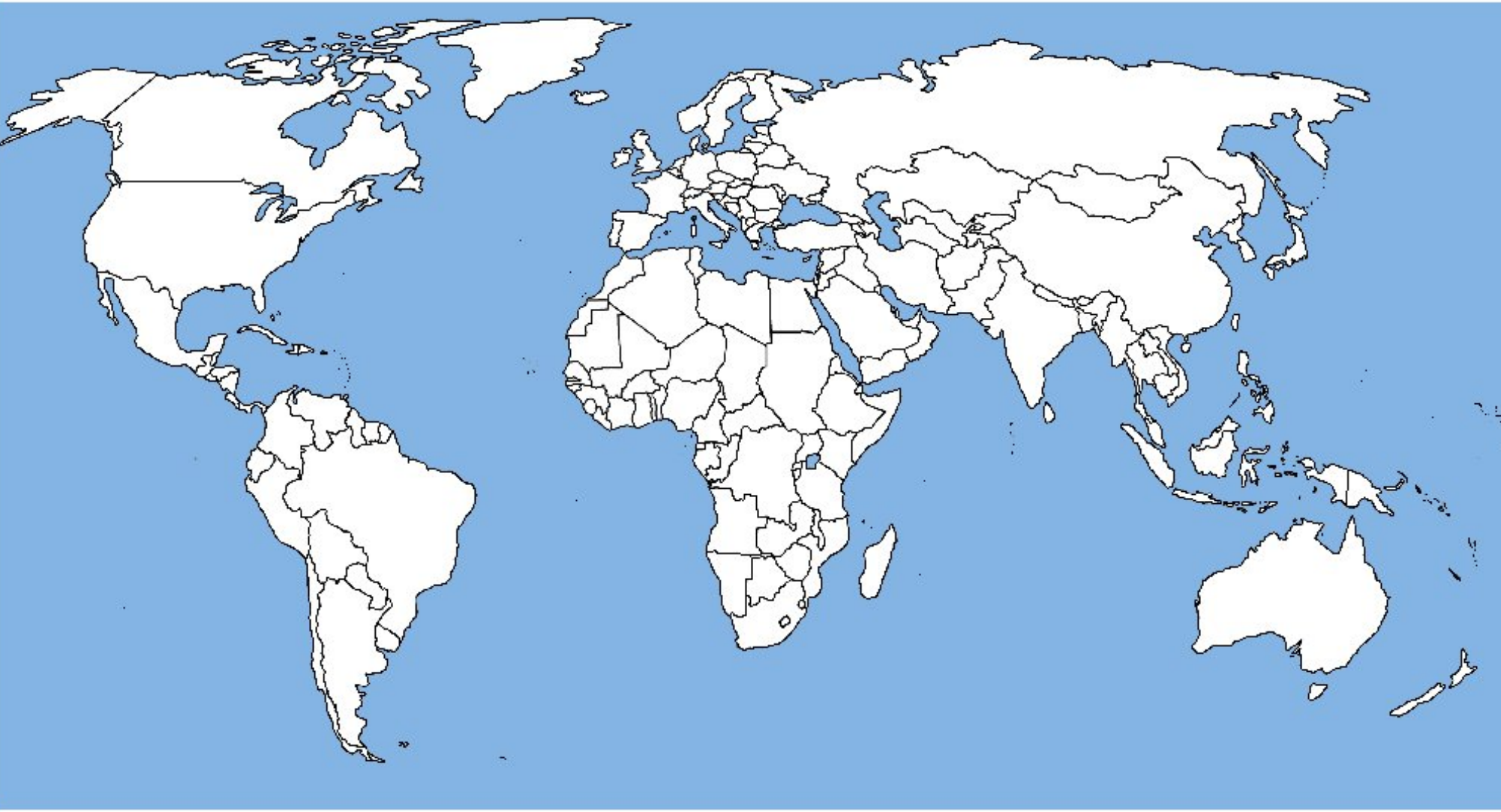
- Domain adaptation, annotation schema, and model combination are all valuable avenues to pursue
- Looking at the effect of Constituent parsers when the language being parsed is not English and the test data is not constituent based
- A new direction might also be effective

A New look

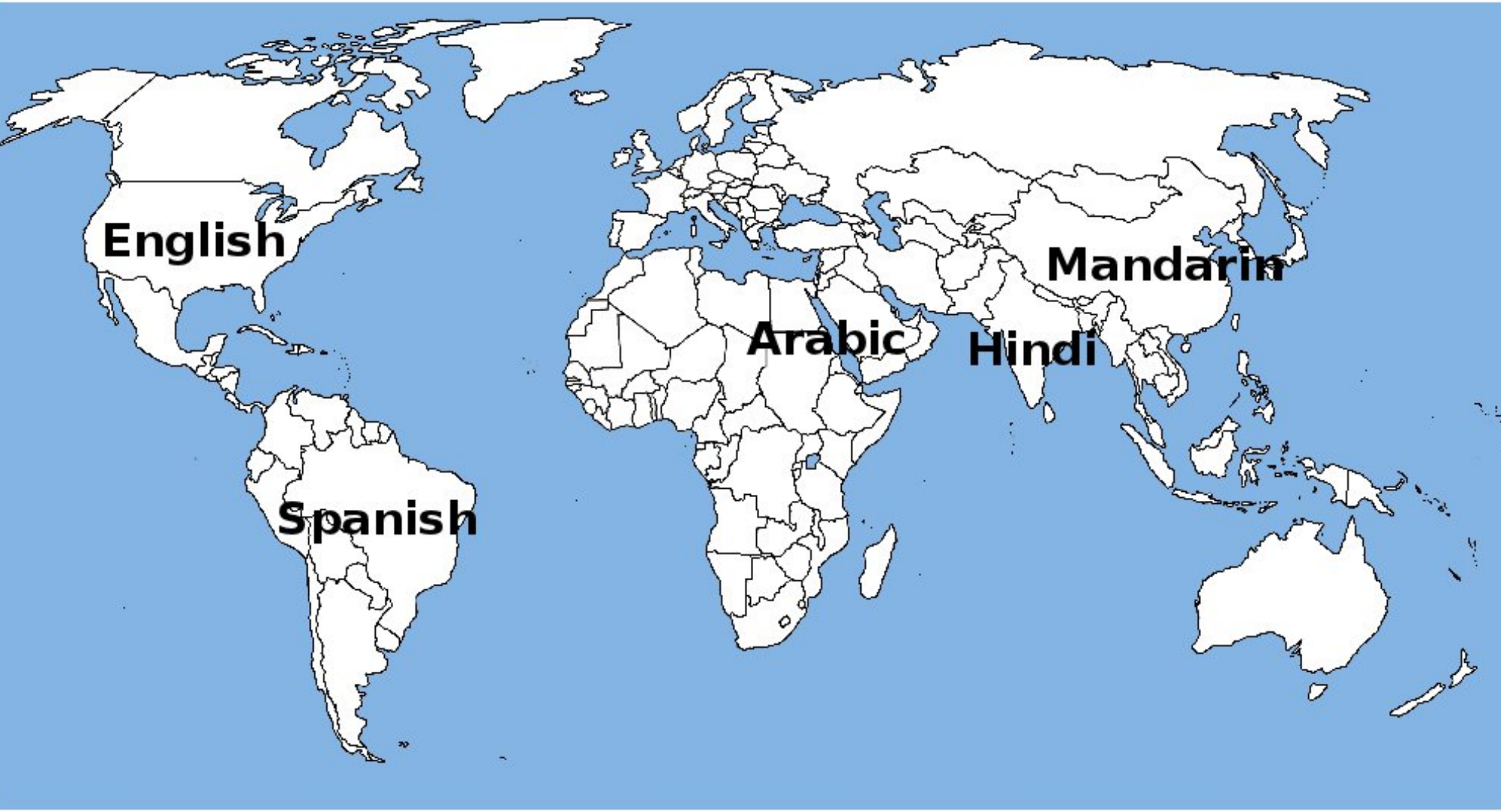
Not just improve accuracy but we want to spread dependency parsing to more languages

- Current methods for unsupervised dependency parsing have extremely low accuracy scores
- It is a worthy goal to try to improve

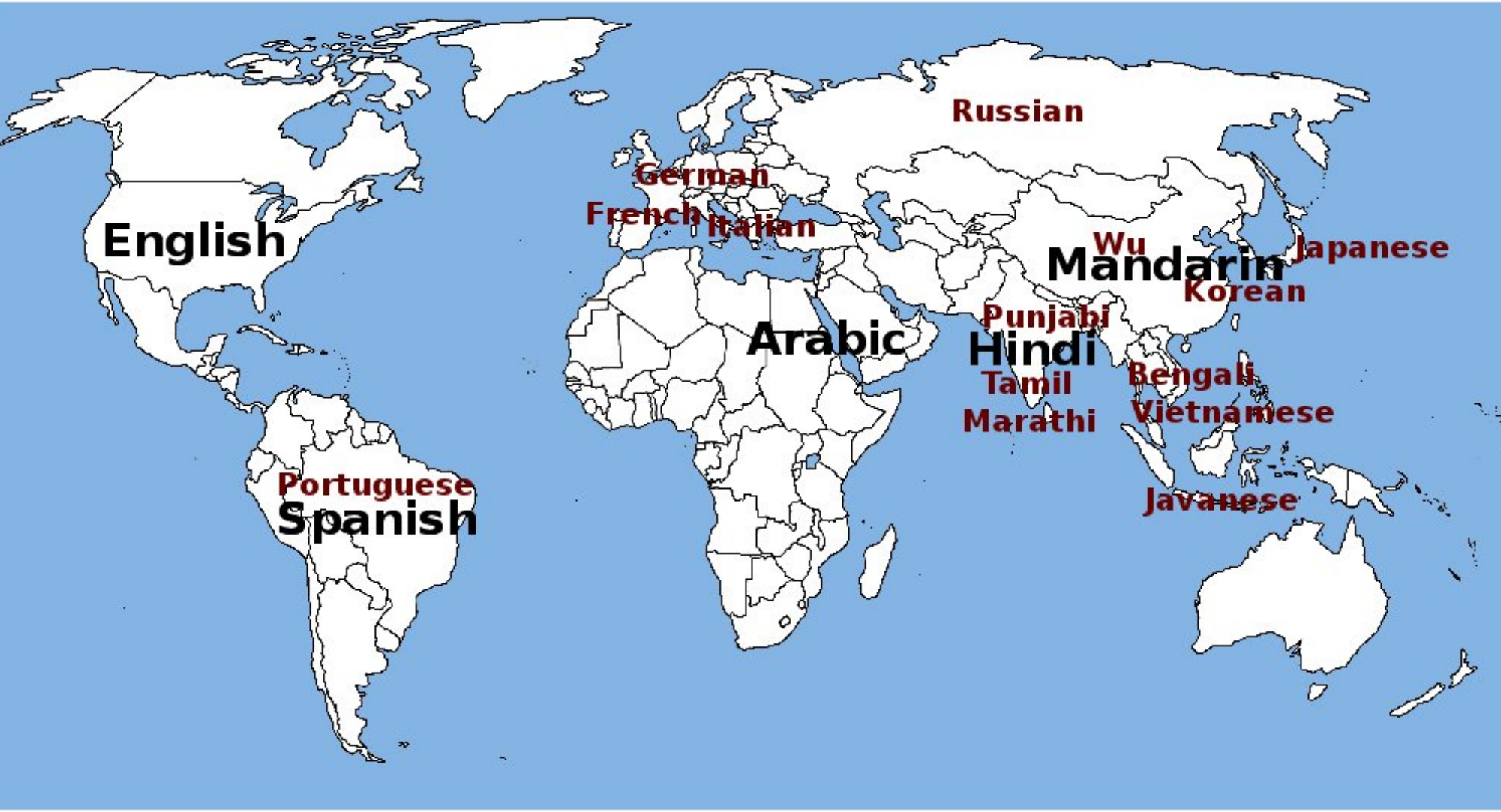
Under-resourced languages



Top 5 Spoken Languages



Top 20 Spoken Languages



And all the others ...

Houston we have a problem

What is available?

- Start with the basics:
  - Word forms: Always available but for under-resourced languages, inherently sparse.
  - Part-of-speech: Some annotation required but far less than dependency relations.

Can we determine heads and dependencies from a POS/word cluster language model?

- Leafs are inherently less important syntactically to the structure of a sentence
  - Can a language model recognize this?
  - Word forms are too sparse
  - Maybe a level of POS language model that can show dependency and the "importance" of certain words

Conclusion

- We have many different ways to approach Dependency parsing each with its own pros and cons. Overall the results are about the same.
- Improvements can be made in Domain Adaptation and model combination
- Annotation Scheme for dependency structures may have an effect on the overall quality of a parser. Empirical evidence for dependency annotation would be positive for the parsing community.
- Future work may be better framed around increasing the scope of available languages in dependency parsing rather than language specific gains.

Questions?