Natural Language Processing

Introduction

Language Processing

- *Level 1* – Speech sound (*Phonetics & Phonology*)
- *Level 2* – Words & their forms (*Morphology, Lexicon*)
- *Level 3* – Structure of sentences (*Syntax, Parsing*)
- *Level 4* – Meaning of sentences (*Semantics*)
- *Level 5* – Meaning in context & for a purpose (*Pragmatics*)
- *Level 6* – Connected sentence processing in a larger body of text (*Discourse*)

Examples of Levels

- L1 : sound
- L2 : Dog - Dog(*s*), Dog(*ged*)
  Lady – Lad(*ies*)
  Should we store all forms of words in the lexicon?
- L3 : Ram goes to market ***(right)***
  goes Ram to the market ***(wrong)***
- L4 : translation from unstructured to structured representation

  go : (event)
  agent : Ram
  source : ?
  destination : market

Example (Contd.)

- L5 : User situation & context

  "*Is that water?*" – the action to be performed is different in a chemistry lab and on a dining table.
- L6 : Backward & forward references –
- Coreference resolution

  "*The man went near the dog. It bit him.*"

  Often co reference & ambiguity go together as in –

  "*The dog went near the cat. It bit it.*"

Approaches to NLP

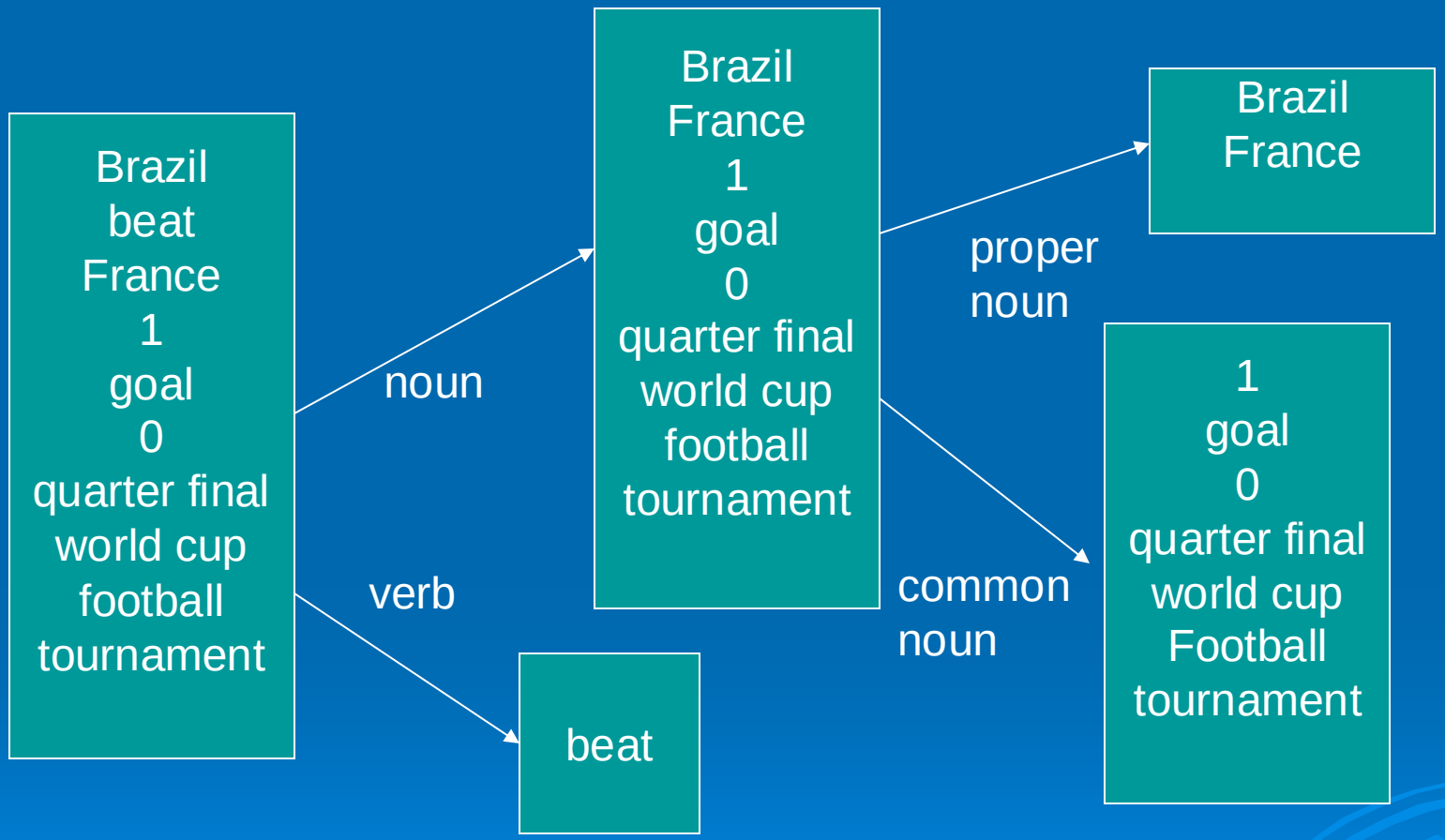
- Classical Approach
  - Human Driven Approach
- Machine Learning/Statistical Approach
  - Data Driven Approach
- Hybrid Approach
  - Partially ML and Partially Classical Approach

ML Approach

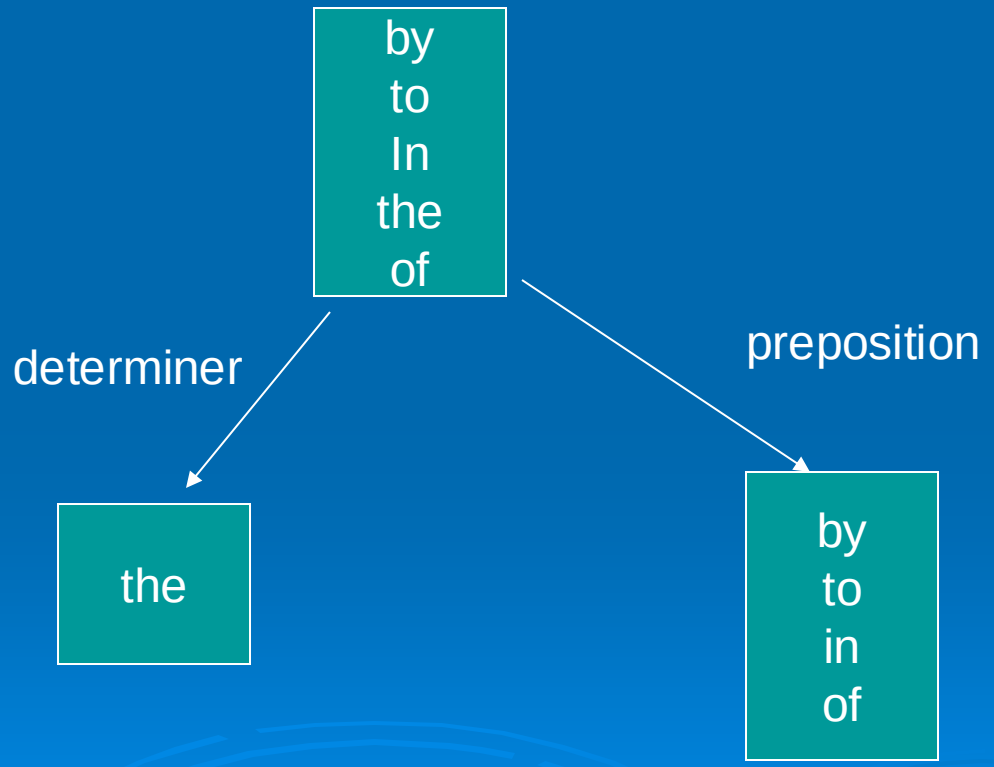
- France *beat Brazil by 1 goal to 0 in the quarter-final of the world cup football tournament. (English)*
- *braazil ne phraans ko vishwa kap phutbal spardhaa ke kwaartaar phaainal me 1-0 gol ke baraabarii se haraayaa. (Hindi)*
- ब्राज़ील ने फ्राँस को विश्व कप फुटबॉल स्पर्दधा के क्वाटर फाइनल में १-० गोल के बराबरी सें हराया (हिन्दी)

Categories of the Words in the Sentence

Further Classification



Further Classification



Why all this?

- information need
  - *who did what*
  - *to whom*
  - *by what*
  - *when*
  - *where*
  - *in what manner*

How is this different from classical NLP

- The burden is on the **data** as opposed to the **human.**