

Lec 22

Linguistics & Comp Linguistics

Linguistics : structures and patterns universals to human langs

- Good for :
- understanding datasets
 - evaluating models
 - analysis
 - ⋮

Sound & Gesture

- Produce / Perceive / Analyse
- Phone ≠ spelling

— deep orthography — shallow orthography — APA —>

- Phonetics — consonants, vowels, voicing, ...
physical property of sounds
- Phonology — phonemes, perceptually distinct units
environment involved

Morphology

- Free / Bound morphemes
- Inflections
- Word formation

Analysis

- Finite state transduction
 - bus-PL → bus-s
 - bus-s → buses
- Just use NN
 - ↳ better at handling unseen input

* Still a good idea to do these analysis for some langs, things like inflexion breaks many English modelling assumptions

Syntax

- POS
- Open class (can easily add things to the class)
vs closed class (those that tend to be fixed)
- Phrases
- Constituents
- Syntax trees
- Dependency trees

* Still good for low-resource langs

Meaning & Intent

- Lexical semantics
- sense — a distinct meaning
 - polysemy — one lexeme multiple meanings
 - ↳ homonyms

→ WordNet

→ Distributional Hypothesis : similar distribution similar meaning

- Compositionality
- new sentences / entities
 - exceptions : idioms / figurative langs

Entailment : given premiss, whether a hypothesis is true

Pragmatics : lang use in context

Speech act theory : meaning consist of intended effect

Presupposition : implicitly assumed context
Which linguist invented the lightbulb?

Implicature : meaning implied, not necessarily literally

Gricean Maxims

Information Structure — how to choose words, syntax, style, etc.

Rational Speech Acts — Bayesian model
↳ communication : recursively thinking what the other person is thinking