NLP1 - Solution Notes 2002

p7 q11
EJB

a) Morph analysis & PoS tagging of words required to deal with variants and to help abstract away from large number of potential words in an open-domain system - statistical approx. wide-coverage & robust (section 4)

Parsing of questions required to identify type of answer (yes/no/wh) & how questioned element relates to rest (debt dir. obj of leave, Quintex subj of leave) - section 5

Parser will need to be wide coverage but also potentially need to know more than PoS of words (e.g. leave as transitive verb). Statistical parsing to deal with ambiguity robustly - section 8

Sentences of documents with matching words (or variants) will need parsing to identify whether they contain appropriate answers using similar representation to facilitate partial matching

Certain questions may contain ambiguities (eg scope, wh-attachment) that can only be dealt with by doing a full parse & semantic interpretation which is difficult in an open-domain → sections 5/6 & 7

Lexical variation may be semantic as well as morphological (debt - debtor - owing - outstanding loans, etc) - section 7

In documents discourse context is important, resolving anaphora (The Fulwell) performing inferences (collapsed ⇒ leave) and possibly word sense disambig (leave a debtor) - section 9

b) Not having much information about words and their relevant syntax & semantics (sections 5/7 & 8) and thus having to resolve indeterminacy statistically

c) Wh-questioned constituent provides clues (who – NP animate entity; where – place name, PP locative etc) but degree of context required difficult to gauge without sophisticated inference ('about/around' = approx amounts, etc) – section 5 & 9

d) Ability to infer that 'collapsed' = bankruptcy & firm only 'leave debts' when bankrupt – virtually impossible in open domain system