P7 q11
EJB

Natural Language Processing 2001

(Paper 7/11)

1) Define TWO of the following four types of ambiguity, giving examples. For each of these two types, describe ONE possible technique for resolving such ambiguities.

   a) Anaphoric coreference ambiguity
   b) Speech act identification ambiguity
   c) Part-of-speech assignment ambiguity
   d) Prepositional phrase attachment ambiguity

[10 marks each]

(Paper 9)

2) The following narrative exemplifies some problems for natural language processing systems. Identify these problems [6 marks] and describe the stages of analysis that would be required to interpret the narrative [6 marks]. Evaluate how successful each stage is likely to be [8 marks].

   Kim ran up a hill by the steepest path. Sandy had egged her on.

Solution Notes

1a) Coreference between a pronoun or definite description and an antecedent is ambiguous when there is more than one potential antecedent that agrees in person, number and/or gender and is a structurally appropriate position, e.g. The cat sat on the table. It was very beautiful (it=cat/table?) Mechanisms for resolution include evaluation of discourse structure (table=obj so more likely to be focus of discourse, etc), poss. with a statistical component and/or use of background knowledge / abductive inference (are cats/table usu. beautiful?)

1b) Speech acts are often indicated by syntactic mood (decl. = assertion, interrog. = question, etc) but not always – indirect sp. acts (e.g. got a light?; can you pass the salt?) Resolution requires arbitrary world knowledge but effective systems can be built in terms of e.g. plan recognition for circumscribed domains (Trains etc)

1c) PoS of open class words (n,v,a) is often ambiguous (storm = n,v; can = n,v, modal, etc) These ambiguities can be effectively resolved by very local context in most cases (I can can a can), but lexical ifromation is critical (can is most frequently a modal)

1d) PP attachment is a common worst-case (exponential) form of ambiguity e.g. I saw the man in the park with the telescope by the monument. Though chart parsers can encode the tree-structured possible interpretations efficiently resolving these

ambiguities require a) general knowledge (telescopes are instruments of seeing etc) or b) statistical preferences that may be partly structurally mediated but are largely lexical (see+telescope is more common than man+telescope)

2)

Problems: ran up = vb+prt or vb+prep; byPP is adjectival/adverbial, definite description with superlative picks out a unique path on the hill via a bridging inference; had cues a discourse explanation rather than narrative sequence; her corefers with Kim (not Sandy) by no reference within clause rule for non-refl pronouns and because Sandy is discourse topic; egged on is an idiomatic phrasal verb discontinuous because of particle movement.

Stages: morph, syn, sem, prag/discourse and descrip of what's involved and potentially resolved at each stage

Eval: morph/PoS tagging shld identify PoS reliably but might treat ran+up incorrectly as vb+prt as it's common (ran up the bill) and might get egg+on wrong as discontinuous; syn might get byPP attachment wrong cos (hill by the lake, location by location, plausible) given knowledge-based or statistical interp.; bridging inference needs deeper world knowledge (hill has inclined paths etc); egg+on is an idiomatic non-compositional predicate, but egg alone is an unlikely verbal predicate, so good semantic lexical info shld resolve this; had is a good simple lexical cue for explanation role / reversed temporal sequencing, so don't need to know about the causal relations between egg+on and run; her is resolved by syn and simple discourse structural rules, so no need for complex world knowledge.