These solution notes are as submitted in January 2003

Natural Language Processing 2003 Paper 9 Question 13 (AAC)

Corresponds to lecture 5 on the syllabus, plus some points in lecture 8 (for part b).

(*a*) (*i*) a feature structure is a connected, singly-rooted, directed acyclic graph, with arcs labelled by features and atomic values on some of the terminal nodes. Labels on arcs leading from a node must be unique.

(*ii*) FS1 subsumes FS2 iff for every path P (sequence of features) in FS1 there is a path P in FS2 and if P has value t in FS1 it also has value t in FS2. If two paths P and Q lead to the same node in FS1 then P and Q also lead to the same node in FS2. Subsumption of FSs gives a partial order.

(*iii*) The unification of two FSs, FS1 and FS2, is the most general structure which is subsumed by both FS1 and FS2, if it exists.

(*b*) Key points:

- CFGs are not expressive enough to represent some phenomena in natural language without considerable redundancy
- examples are agreement, subcategorization, long-distance dependencies
- (not discussed in detail in lectures, but the answer could point out CFGs are not expressive enough to represent some phenomena in some natural languages at all, unless one assumes that there are bounds on recursion)
- CFGs are more efficient than FS grammars – FS copying, unification of large structures, packing.
- Probabilistic CFGs can be defined.
- CFGs can be automatically converted to FSA for speech recognition.
- CFGs are good for spoken dialogue systems: limited domain, control speech recognition.