

Comments Please see the attached report.

---

Comments on the examination

1. General comments

**These comments are made in relation to unapproved scores.**

63 candidates sat the examination. The average mark was 57.6%, up from last year. The top mark was 85%, the bottom mark was 29%. 14 candidates scored under 50%. 9 marks were in the distinction range and 20 in the merit range. All but 1 candidate answered question 1 (av. ~60%); 47 candidates answered question 2 (av. ~57%); 21 answered question 3 (av. ~41%); 51 answered question 4 (av. 64.5%); and 8 answered question 5 (av. ~33%). In several cases, parts were not attempted, which would clearly lead to depressed scores. Also, there were several vague, general, over-brief answers that did not persuade that the candidates really understood points at issue. Marks were lost by some, due to not answering all (especially discussion) parts of questions. Several candidates overanswered (including by ignoring instructions that would have saved them time).

2. Specific comments (not all parts merited comment)

*Question 1*

a) Several candidates could not classify appropriately and apparently were not able to differentiate a syntactic (structural) ambiguity from a lexical ambiguity. Several gave lengthy explanations that were not called for, given the marks available (and in some cases unfortunately revealed that the basis of their otherwise correct classification was wrong).

b) i) Many answers failed to take account of the difference between training a tagger and applying a tagger. The question concerned applying a tagger: "*used to tag a text*". One does not do training at the same time as tagging, as many answers tried to convince me. It is important to read a question to understand what is being asked.

c) i) Several answers explained how a PCFG could be used, but did not explain how its rules are derived. Again, it is important to understand what is being asked.

d) ii) was answered poorly by several as they drew a tree or several subtrees with no further comment, i.e., no explanation of steps. Those that referred to the grammar and explained how a naïve top down depth-first parse would deal with backtracking did well. Some candidates drew structures that had no corresponding rule in the grammar. Some candidates attempted to explain how the sentence "many fish flooded rivers" would be parsed, but this was not the sentence of this question part.

iii) Generally well answered, but with evidence of brain-dumping, particularly by those who had done less well in d (ii): they were able to list backtracking as an issue for a top down parser, *but had not been able to use that knowledge in answering d (ii)*.

iv) Variably answered, with several candidates failing to note the need for extra predictions on [0,0] and also further predictions on [1,1]. Some gave edges all the way through the chart

and ended up with far more edges than the 9 asked for. This was simply wasting valuable time as I had not asked for a complete parse. Some did not indicate what the start and end positions of edges were.

### *Question 2*

a) Some glossed BILOU but did not actually say how this should be interpreted (i.e., what does e.g., 'Last' mean?). Many did not explain the relevance to machine learning methods.

b) Variably answered. Many specified e.g., B-PERSON, I-ORG, etc., whereas the question called for noun phrase chunks, i.e., B-NP, I-NP, etc. Several only attached BILOU tags to the underlined items, whereas the instruction was "*Annotate the following sentence*". Several did not attach any BILOU tag to punctuation. NB: each token carries a BILOU tag.

d) This was underanswered by several, with only a very few examples or types of issue being addressed.

e) i) and ii) These were not well answered, with many giving the same features largely for detection as for classification.

e) iii) Some underanswering occurred here too, with also some not addressing all three approaches.

f) Several candidates thought it appropriate, having pipelined several components to add annotation to enrich the original document, to then remove all annotations at a later stage, a somewhat strange decision. Some attempted gazetteer lookup without any tokens available to be looked up. Some attempted to apply a syntactic parser without there being any sentence, tokens or POS-tagged data as input for it. Some thought that evaluation could be done with one input, despite the reference evaluator being described as taking two inputs. Some had their first analysis component attempting to analyse with no input available.

### *Question 3*

a) Some candidates generated orphans (a standalone link between two words, but nothing else linking in to the overall graph). Others had wrong direction for a relation. Some had wrong labels or wrong dependencies.

b) i) The scope of PERSON caused problems for some. Some strange decisions were made, e.g., to annotate National Nature Reserve as a NamedOccasion.

b) ii) Many answered that "May" filled the criterion. Unfortunately, "May" is not a sequence of tokens (a NER labels a single token or a sequence of tokens as a NE, the question concerned sequences).

b) iii) Some very good answers and some very poor ones here, the latter revealing a lack of knowledge about events and event participants. Some ignored the instruction not to reproduce the role definitions, thus wasted valuable time. Some applied templates in the order they occurred. Some considered "London's Wembley Stadium" to be a Place-Arg, despite the data being "London's iconic Wembley Stadium" (which should lead to a different analysis of NEs than the one assumed).

c) iii) There were several good answers although many underanswered for the marks available. Some did not consider how search would or could operate within a semantic search perspective, or how NER or event extraction could support the creation of faceted search.

#### *Question 4*

Several only answered the statistical parts of this question.

a) Many candidates did not refer to WN relationships beyond synsets. Some thought that "boy" and "girl" were antonyms and that "coat" and "coat button" were synonyms. Thought is required: wearing just a coat button would not keep you very warm compared to wearing a coat. Several characterised relationships properly but did not say how an analyser could use them (WN is not an analyser, it is a resource).

c) i) Some only discussed two of the three approaches.

c) ii) Some tried to convince me that BAM and BMM could construct a thesaurus. BAM and BMM both output vectors. Further steps are necessary to take these vectors and construct a thesaurus.

e) i) Some did not state the *t test* null hypothesis, others stated it only in general terms and not for the data under analysis.

e) iv) Calculations were in general well tackled. Credit was given for part work and for manipulation of formulae even though an error may have been introduced that prevented arriving at the perfect result. In several answers, a value that had been calculated correctly earlier and assigned to a variable was substituted with a different value in a later calculation using that variable, credit was given. Credit was also given in cases where e.g., a scientific notation that was being correctly calculated and used (as in  $5 \times 10^{-5}$ ) was later expressed with the wrong number of zeroes. In general, if solid understanding was being shown, this was given credit. Some candidates did not calculate a result for their final expression: as all inputs to that expression were correct, due credit was given. Some candidates calculated also PMI. This was a waste of time as no marks could be awarded for this.

#### *Question 5*

Only 1 candidate of the 8 who answered this question obtained a pass mark for the question. Several did not answer all parts.

a) There were some confusions over working out a result for kappa.

d) i) and ii) were rather poorly answered, with a great deal of underanswering given the 9 marks in total available for these parts.

e) There were some over-brief (given the marks available) and general answers to this discussion-type question, and these did not appear to be well-related to aspects of the unit. Hardly any answer brought in evidence from the literature (i.e., from reading around), which would indicate an over-reliance on lecture slides.