Comments on the examination

1. General comments

These comments are made in relation to unapproved scores.

62 candidates sat the examination. The average mark was 60.72%. The top mark was 90%, the bottom mark was 5%. 11 candidates scored under 50%. 20 marks were in the distinction range and 16 in the merit range. 57 candidates answered question 1; 53 candidates answered question 2; 54 answered question 3; 3 answered question 4; and 18 answered question 5. 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.

2. Specific comments

Question 1
a) This is a bookwork question yet many did not get full marks mostly due to missing the garden path type of ambiguity.

b) i) The majority answered this question satisfactorily. However, several candidates did not follow the instruction to annotate ALL tokens in the sentence. Also, some annotated for part of speech or for named entity, when the task was to annotate for noun phrase chunks. Lesson: read the question, do not make false assumptions.

b) ii) In general, this question was not addressed well. Many stated that performance depends on the data but only very few elaborated on their answer or mentioned the possibility of sparsity, e.g., that the tokens that BILOU specifies better could be sparse in the dataset (tagset too large in relation to the dataset).

b) iii) Some lost marks here in not providing fuller detail, particularly about implications for processing and (re-)constructing parse trees.

c) Some candidates incorrectly (and fatally) inserted the annotation remover towards the end of their workflow: once linguistic annotation is removed, any following component that depends on such annotation has no input to work with. Others had a branch that started with the sentence segmenter, but had no input for it to work on. Others invented a document reader component: this was not in the given list, however, but the gold standard corpus reader was. Some produced parallel workflows for the 3 recognisers, however this revealed a misunderstanding of the role of the CAS (components accept a CAS and add to it if they can, and pass it on to the next component). Only the reference evaluator could accept two inputs. Some ordered POS tagger before the tokeniser, giving it no tokens to work with. Some decided to include the syntactic parser and ordered it before the POS tagger, giving it no POS to work with. Due attention must be paid to the nature of information that some component expects as well as what it produces.
d) This was rather poorly answered by most, and in very brief and general terms by many.

e) Some did not answer this question part at all. Many made good attempts at producing a structured argument to make motivated recommendations.

Question 2

a) This was generally well tackled. Some gave no discussion at all or did not indicate what impact decisions might have on following components. Some claimed that forms such as BBC would present a problem for a tokenizer, which is not true. However, the form B.B.C. would present a problem due to the dots. Some claimed that tokenizers would have problems recognizing named entities consisting of several tokens: this is not the job of a tokenizer, but of a NER.

b) i) Several answers showed there was confusion over training a TBL tagger and applying a trained tagger to totally new, unseen text, particularly regarding the guesser component. There are no unknown words in the training text, as it has been derived from the gold standard corpus. Therefore, a guesser is not needed during training, only in tagging totally new, unseen text. Most answers told us that rules were induced, but did not think to tell us what these rules looked like or how they should be interpreted.

b) ii) Some gave the correct order, but with no justification (so we could not tell if this was just a lucky guess). Others gave some justification but did not explain precisely why lexical rules lead to diminishing returns.

c) i) Majority of the students drew the dependency graph correctly for one interpretation (the man has the binoculars). However, for the other interpretation (the girl is using the binoculars), many students wrongly connected "girl" to the preposition "with", rather than the verb "watched".

c) ii) Several answers showed confusion regarding where to connect the phrase "with the binoculars". It should be connected to the main verb phrase starting with "watched" rather than the root of the phrase structure tree.

c) iii) Because many students wrongly attached "girl" to the preposition "with", consequently many answered incorrectly when asked to list the predicate-argument structures for one of the possible interpretations (the girl is using the binoculars).

Question 3

a) This question was answered well overall.

b) i) This question was answered well overall.

b) ii) Most students mentioned named entities as a possible type of feature but not syntactic relations (e.g., between the token *run* and entities).

b) iii) Many students defined what facets are but did not explain that text mining helps in developing them by generating annotations (e.g., named entities, events) which can serve as semantic metadata.

c) This was mostly answered well, although several candidates ignored the meaning of NOT (set complement) and took just the posting list size of the term in the scope of the NOT. Others showed they had correctly calculated the sizes but then did not respect the strict order, e.g., "100 AND 90" should be rather "90 AND 100" for processing purposes, even though, theoretically, the set operation is the same for both orders. However, for processing purposes, the "merge" procedure in practice used on the *sorted* posting lists must start with the smaller of the two, to be efficient.

d) Most candidates did well here, although some gave confused or unclear explanations. Some did not distinguish inverse document frequency from document frequency.

Question 4

a) Some candidates did not refer to WN relationships beyond synsets, or discuss disadvantages. Some did not refer to any WN relationships, but to subject-verb-object patterns, incorrectly for a question on lexical relations.

b) i) This was generally well tackled.

b) ii) Some did well here.

c) This was answered well by some, but in general terms by others or the answer remained underdeveloped for the marks available.

d) i) No answer was completely correct.

d) ii) This was answered partially by some, in terms of the first part, but the part on the simplest sort of NN that could be used in this case was not well answered.

Question 5

a) i) The majority of students answered this question well.

a) ii) This question was answered well.

a) iii) This question was answered well.

a) iv) This question was answered well, although a few did not provide sufficient justification on why micro-averaged scores give a better indication of overall performance (as macro-averaged scores are easily affected by the sample size for each named entity type).

b) This was not answered by some, thus sacrificing a quarter of the marks available for the question. Others gave good answers with sound grasp of detail and challenges.

c) There were some over-brief (given the marks available) and general answers to this discussion-type question, and other very good ones that showed very good awareness. Some again did not answer this part, which was disappointing for the examiners who expected any

student following the unit to have views (positive, negative or neutral) on relevant systems and technologies, and on the way the field was likely to develop.