Please see the attached.
These comments are made in relation to unapproved scores.

35 candidates sat the examination. The average mark was 54.62%, slightly down from last year. The top mark was 80%, the bottom mark was 16.66%. 12 candidates scored under 50%. 8 marks were in the distinction range and 8 in the merit range.

All but 2 candidates answered question 1 (av. 55.15%); 20 candidates answered question 2 (av. 59.25%); 16 answered question 3 (av. 63.63%); 23 answered question 4 (av. 50.43%); and 13 answered question 5 (av. 46.15%). 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. One candidate answered more than 3 questions, which is generally a very bad strategy to follow.

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 (and in some cases unfortunately revealed that the basis of their otherwise correct classification was wrong).

b) i) Most answers were reasonably good, although several failed to mention the role of templates.

ii) This part was less well answered, with some believing that lexical rules contribute more.

c) Some candidates attempted to operate with a general language notion of independence and sensitivity, which would indicate a lack of knowledge. Others correctly described the notion of independence assumption but did not explain well the lexical sensitivity issue nor how these disadvantages could be overcome.

d) i) This question turned out to be what examiners call “a good distinguisher”. Most candidates were able to correctly construct one parse tree, however were either unable to construct a second, or constructed one that did not reflect the capability of the grammar (e.g., there is no rule VP → NP VG which some tried to use). A second parse tree can be constructed with this grammar which is a valid tree with respect to the grammar. That tree may not make semantic sense, it may violate agreement, but the grammar is a syntactic grammar, not a semantic one, and it does not have a useful
mechanism to handle agreement (some grammars do, this one does not). All that is called for here is to apply rules that validly apply, and doing so will lead to two parse trees for the sentence.

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 top down depth-first parse would deal with backtracking did well, as did those who adopted a top down breadth-first strategy.

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 time.

Question 2

a) Some did not explain U or L correctly. Some did not give reasons for preferring BILOU.

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. Several did not attach any BILOU tag to punctuation. NB: each token carries a BILOU tag. (If someone had said, e.g.: I assume the token is “countries,” and given it a tag, I would have accepted that.)

c) Some addressed only 2 of the 3 types of annotation.

d) Some did not make clear what their informal notation meant (e.g., whether two items were in sequence or were members of a set (alternatives)). Several specified very general patterns that would incorrectly match other elements in the text. Others did not notice the need to take care of a person name element starting with lower case. Some were unclear about scope of the entity matched, e.g., is a title part of a person entity? (or is it an attribute of the entity?).

e) ii) Poorly answered by several, who were convinced of the sufficiency of gazetteer lookup for NER, or who proposed adding further entries to the gazetteer. This question part was based on lab and coursework thus one might have expected a firmer understanding.

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.
c) ii) Some very good answers and some very poor ones here, the latter revealing a lack of knowledge about events and event participants.

d) Several answers described the type of facetted search available on e-commerce Web sites, which bore little relation to topics of the course, and did not consider how facetted search would or could operate within a semantic search perspective, or how NER could support the creation of facetted search.

e) Some good attempts, although in general many did not recognize the role of inheritance in a UIMA type system.

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.

**Question 4**

a) Some candidates referred to characteristics that are in fact common to many dictionaries, not to those that make WordNet unique.

c) Some focused on frequency of occurrence of words in general, without considering the distributional aspect.

d) Some took a different interpretation of compositional. I accepted this, although it would indicate that there was a lack of knowledge about compositional methods in the distributional semantics framework.

e) 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. This was taken as a slip and full credit 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 have a calculator, but were able to demonstrate in their working that they could arrive almost to the end, but could 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**

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

c) i) Several candidates did not pick up on the phrase “text elements” and gave therefore inappropriate answers.

ii) Some gave only descriptions without addressing types of technique or resources.
d) 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.