

Comments COMP24412 exam review

These comments are primarily for the second marker, and for the external examiner and anyone else involved in investigating the issues relating to Q2(c), but I have no objection to them being made available to students at the exam review session along with the original model answers.

The exams officer and the director of undergraduate studies have both seen this document and have agreed that my proposal for dealing with the problems arising from Q2(c) is reasonable, if a touch generous.

Overall

This paper had an error on one of the questions. The question was about chart parsing: students were asked to show the steps that a chart parser would take when analysing the sentence I know he ate it using a grammar and a lexicon which were aimed at the previous part of the question, which asked them to demonstrate top-down or bottom-up parsing on the sentence He runs a shop. This question had 8 marks out of 100 allocated to it, suggesting that it should have taken about 8% of the time for the exam, i.e. about 9.6 minutes. I think that this was a reasonable estimate of the time that this question part would have taken.

Almost every student answered this question (91 out of 99). What people did can be summarised as below:

- (i) A small number of students just said "This isn't do-able with the grammar and lexicon provided" and skipped the question.
- (ii) Some tried to do it using He runs a shop as the target sentence, with the grammar and lexicon provided in the paper. This was what I had announced as the right thing to do when the problem was pointed out to me.
- (iii) Some tried to do it by improvising a grammar that would account for I know he ate it.
- (iv) Some did that, and then re-did the question using He runs a shop following the correction (nearly everyone who did this did it right both ways).

For anyone who did (ii)-(iv), I marked what they gave me in terms of whether it was a sensible trace of a chart parser, and hence demonstrated that they understood how chart parsers work, irrespective of what grammar they used and which sentence they used as the target. I was reasonably, but not excessively, generous in my marking of these questions. I can't really see that people who did (ii) or (iii) were substantially disadvantaged by the mistake, and if everyone had done that I would be inclined to leave the marks as they stand.

But not everyone did that. Some people did (iv), and these people clearly lost time. So the fair thing to do for them would be to compensate them for the lost time, e.g. by multiplying their mark by 1.087, on the grounds that that they lost 8% of the time for the exam, and hence could conceivably have got 100/92 times their actual mark if they had not wasted time redoing this question. I think that one of these people was the person who got 90% overall, including 34/35 for this question, but multiplying their mark by 1.087 would not take them over 100%, so at least we don't run into that problem. And some people did (i), which wasn't unreasonable, so they only did an exam with a maximum of 92 marks, so their mark should also be multiplied by $100/92 = 1.087$.

So that says that people who did question 2 and who did either (i) or (iv) should have their marks multiplied by 1.087. What about people who did (ii) or (iii)? They may have lost time, or been distracted by the problem, and if it were clear that this had happened then they should be treated as though they had done (iv) (lost time) or (i) (been distracted), and have their mark multiplied by accordingly. But it's not possible to tell what effect the issue had on them, so we can't multiply some by 1.087 and not others. So I think that that says that we have to multiply their marks by 1.087.

And then we have the small set of people who didn't do Q2 at all. They could argue that the reason that they chose not to do Q2 was because they spotted the problem and decided to steer clear of it. It would be an entirely implausible argument, because you wouldn't actually spot it until you were pretty near the end of the question, but it's not disprovable. So they could argue that they were disadvantaged because they had been forced to choose a question that they wouldn't otherwise have chosen. Multiplying by 1.087 seems like as reasonable a response to this as anything else.

The overall average for the exam without scaling it in this way is 55%. Multiplying that by 1.87 takes it to 59%, which is the suggested target outcome. I would not do this scaling were it not for the issue with 2c, because I am fairly certain that the coursework will bring the overall average up substantially anyway, which is how I plan it: coursework which is primarily intended as a learning exercise, and which you get quite high marks for doing, and then an exam which is more challenging and which checks that you did indeed do and learn from the coursework, rather than just getting someone else's answer and submitting it. So I would not in general think about scaling the marks for this course until I had the combined mark.

I did some scattergrams and things to see if I could see any patterns, but nothing emerged. For each question, some people made a mess of it and some people did it just about perfectly. As I was marking I got the impression that people who couldn't do the Prolog question were also consistently doing badly on other questions, but nothing obvious of that kind showed up when I looked for patterns (and I don't know what that would have told us anyway, except that people who didn't do the Prolog labs probably didn't do the other labs either, which wouldn't have been a great surprise).

More local issues:

Q1(d) Several people said that [A, B, C] can't be unified with [[A, B, C], T]. The Prolog that they were using doesn't do an occurs check by default, so I was expecting an answer that involved a circular structure, but if they said that this couldn't work I accepted it. BUT if they argued for that, then I did not accept an answer for part (ii) that unified A with `syntax(head(cat(noun), A, B, C), D, args([]))`

Q2(a) Some people answered this question by assuming that the CFG would only have very simple categories, and hence that you couldn't make the required distinction. Some people did what I wanted, and said that you'd have to write a very verbose and fiddly CFG if you wanted to deal with this issue. I did give full marks to people who took the simple view, but I did demand that answers contained a reasonable level of detail, preferably including a sample grammar. I think that was fair enough, since the question did include the words "you should illustrate [your] answer by considering how you [would] write a grammar that ..."

Q3(b): the definition of compositionality that I gave in the notes was poorly worded. I gave 2/2 for people who repeated my flaky definition: the most important part of the question was, in any case, in the working out of the example in part (c). The mark for

What problems do each of these forms of ambiguity pose for hopes of providing a compositional treatment of semantics? [5 marks]

was probably a bit generous. I marked this bit by distributing the 8 marks for

Explain what lexical ambiguity, structural ambiguity and scope ambiguity are. You should illustrate your answer with examples of each kind of ambiguity. [3 marks] What problems do each of these forms of ambiguity pose for hopes of providing a compositional treatment of semantics? [5 marks]

over whatever they wrote for this section.

Q4c: I did stick to the statement in the model answer that "I absolutely need them to collect the defaults and attempt the proof of absurd at the end". Just telling me that you can't prove `white(bruce)` & `black(bruce)` because they are inconsistent wasn't enough -- I needed to know that you have to do a consistency check at the end of the other proofs as well.
