Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Natural Language Systems

Date: Wednesday 24th May 2017

Time: 09:45 - 11:45

Please answer any TWO Questions from the THREE Questions provided

Use a SEPARATE answerbook for each QUESTION

This is a CLOSED book examination

The use of electronic calculators is permitted provided they
are not programmable and do not store text

Question 1

a)

i) Describe the difference between inflectional and derivational morphology. Illustrate your answer with reference to the words '*reconstructions'* and '*unreconstructed*'.

[4 marks]

ii) Which of these (inflectional or derivational) would you be more concerned with if you were developing a machine translation system? Explain your answer.

[4 marks]

b) Consider this grammar:

S → NP VP	ADJ → *complex*
N → DET NOM	ADJ → *first-year*
NOM → ADJ NOM	NOUN → *complex*
NOM → NOUN	NOUN → *first-year*
VP → VERB NP	NOUN → *houses*
VP → VERB	NOUN → *students*
DET → the	VERB → *houses*

i) Show how this grammar can be used to derive a parse tree(s) for the sentence '*The complex houses first-year students.*'

[2 marks]

ii) Explain what kind of ambiguity the word '*complex'* exhibits in the sentence above.

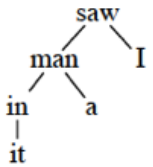
[2 marks]

c)

i) Explain the data structures (stack, queue) and operations (shift, reduce left, reduce right) used in transition-based dependency parsing.

[4 marks]

ii) Show the sequence of operations that would obtain the dependency tree below from the sentence '*I saw a man in it*'.



[4 marks]

[Question 1 continues on the following page]

[Question 1 continues from the previous page]

d) Explain how you might use the vector space machinery for calculating the similarity between documents in order to choose between different WordNet synsets for a given target word. Discuss any issues that should be considered.

[8 marks]

e) You have been asked to develop a named-entity recognition (NER) system to identify mentions of drugs and medications appearing in social media posts.

i) Explain the main challenges that you may face when designing your NER system in this context.

[3 marks]

ii) Discuss advantages and disadvantages of rule-based and machine-learning approaches to NER.

[6 marks]

iii) Your system is part of a wider project that aims to identify drug side effects. Would you aim for your NER system to have higher precision or higher recall? Explain your answer.

[3 marks]

[PTO]

Question 2

a)

i) Explain the roles of tag-transition and word likelihood probabilities when using an HMM (hidden Markov model) for part-of-speech (POS) tagging.

[4 marks]

ii) Describe how these probabilities can be obtained using a POS-labelled corpus.

[4 marks]

iii) Given the sentence "*I promise to back the bill.*" and tables of tag-transition and word likelihood probabilities below (obtained from a POS-labelled corpus) show how you would compute the probability of "*back*" as a verb (VB) versus the probability of "*back*" as a noun (NN). Assume that the tagger has already tagged all the previous words and that the tables below contain all tags applicable here.

	VB	TO	NN	PPSS
<s>	.019	.0043	.041	.067
VB	.0038	.035	.047	.0070
TO	.83	0	.00047	0
NN	.0040	.016	.087	.0045
PPSS	.23	.00079	.0012	.00014

Tag transition probabilities. The rows are labeled with the conditioning event. Thus, P(VB|<s>) = .019. POS tags: VB = verb, base; TO = infinitival *to*; NN = noun, singular; PPSS = pronoun, not 3rd person singular.

	I	*promise*	*to*	*back*
VB	0	.0093	0	.00008
TO	0	0	.99	0
NN	0	.0085	0	.00068
PPSS	.37	0	0	0

Word likelihoods

[4 marks]

b)

i) Explain the difference between homonymy and polysemy. Give an example of each that illustrates your point.

[4 marks]

ii) In the context of word-sense disambiguation, explain the 'vector space' model for representing lexical meaning and discuss word weights one can use.

[4 marks]

iii) What can constitute the 'context' in such models? What kinds of features can be used to represent contexts?

[4 marks]

[Question 2 continues on the following page]

[Question 2 continues from the previous page]

c) Natural language systems typically include a component that tries to find the syntactic relations between words. Some systems use a hand-crafted grammar, which is intended to capture the constraints that allow a native speaker to decide whether some sequence of words is a well-formed sentence of their language. Data-driven systems attempt to infer these rules from sample annotated data, usually in the form of collections of dependency trees. Discuss the advantages and disadvantages of both approaches.

[6 marks]

d) Suppose you are building an information extraction (IE) system to identify the city and state in which a person was born.

i) Explain what bootstrapping in the IE context is. Describe how you could use, for example, Wikipedia, to find patterns that could be used in general to determine place of birth.

[6 marks]

ii) Would it be better to use, for example, both Google and Wikipedia rather than one corpus alone to find patterns? Explain your answer.

[4 marks]

[PTO]

Question 3

a)

i) Explain the term frequency and inverse document frequency (tf-idf) weighting scheme in the context of information retrieval. Justify the design of inverse document frequency in tf-idf weighting?

[5 marks]

ii) Consider a target document containing 500 words wherein the word '*computer'* appears 5 times. Assume there are 10 million documents and the word '*computer'* appears in one thousand of these. What is the tf-idf weight of the word '*computer'* appearing in the target document?

[3 marks]

b)

i) Describe the bigram model in the context of information retrieval, and explain how it can be used to compute sentence probability.

[5 marks]

ii) Based on your description, calculate the probability of the sentence *"I want to have British lunch."* by utilizing the trained bigram probabilities as provided in Table 1.

[5 marks]

<s> I	0.25	I want	0.32	want to	0.65
<s> I'd	0.06	I would	0.29	want a	0.05
<s> Tell	0.04	I don't	0.08	want some	0.04
<s> I'm	0.02	I haven't	0.04	want thai	0.01
eat Thai	0.03	British food	0.60	to eat	0.26
eat British	0.001	British restaurant	0.15	to have	0.14
eat Chinese	0.02	British cuisine	0.01	to spend	0.09
eat a	0.04	British lunch	0.01	to be	0.02

Table 1. Bigram probabilities trained by "Berkeley Restaurant Project."

c)

i) What is the difference between the way that people use their articulators to produce vowels and the way they use them to produce consonants?

[4 marks]

ii) On the basis of what you have just said about this difference, decide which of the following speech signals and intensity contours (see Figure 1 on the following page) corresponds to someone saying '*kick'* and which to someone saying '*icky'*.

[4 marks]

[Question 3 continues on the following page]

[Question 3 continues from the previous page]

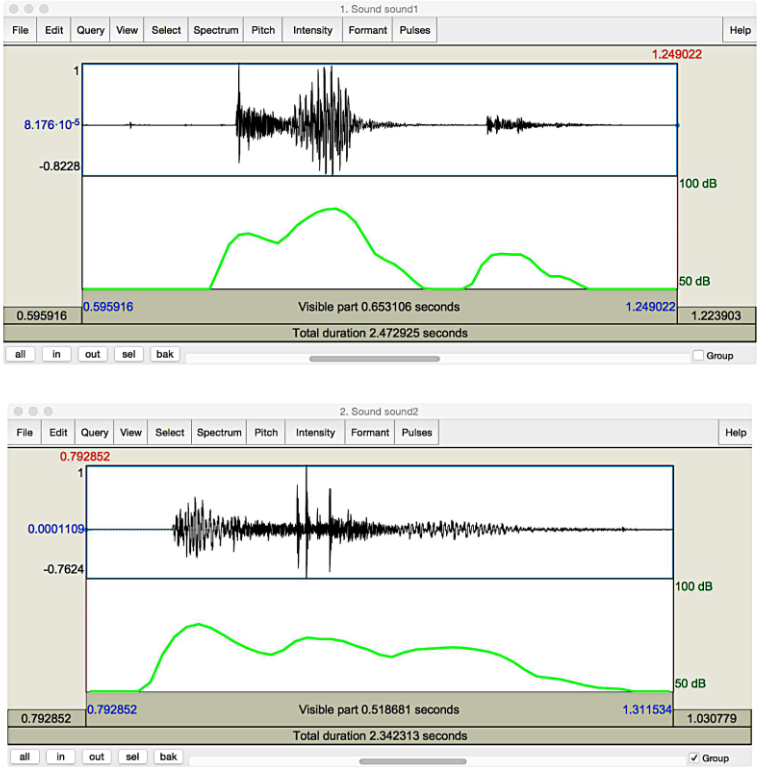


Figure 1. Recordings of ‘kick’ and ‘icky’

d)

i) What is the main drawback of using only pronunciation dictionary in speech synthesis?

[3 marks]

ii) What methods can be used to accompany pronunciation dictionary to overcome such drawback?

[3 marks]

e)

i) A speech recognition system receives the acoustic input and picks the source sentence that best matches the input. What are the two main problems the system is required to solve?

[4 marks]

ii) Denoting an acoustic input as O and a candidate sentence as W, explain why the best matching sentence can be computed by $\hat{W} = \underset{W}{\operatorname{argmax}}\, p(O|W)p(W)$, and state the assumption that is made to obtain this formulation.

[4 marks]

END OF EXAMINATION