DCU

DUBLIN CITY UNIVERSITY

SEMESTER 2 EXAMINATIONS 2013/2014

MODULE: CA446/A – Statistical Machine Translation

PROGRAMME(S):

CASE	BSc in Computer Applications (Sft.Eng.)
MTT	MSc in Translation Technology
ECSA	Study Abroad (Engineering & Computing)
ECSAO	Study Abroad (Engineering & Computing)

YEAR OF STUDY: 1,4,O,X

EXAMINERS:

Dr Jennifer Foster (Ext:5263)
Dr. Ian Pitt

TIME ALLOWED: 2 Hours

INSTRUCTIONS: Answer Question One and two other questions.

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE INSTRUCTED TO DO SO
The use of programmable or text storing calculators is expressly forbidden.
Please note that where a candidate answers more than the required number of questions, the examiner will mark all questions attempted and then select the highest scoring ones.

Requirements for this paper (Please mark (X) as appropriate)

	Log Tables		*Thermodynamic Tables*
	Graph Paper		*Actuarial Tables*
	Dictionaries		*MCQ Only – Do not publish*
	Statistical Tables		*Attached Answer Sheet*

QUESTION 1 ***[TOTAL MARKS: 40]***

Answer EIGHT of the following ten short questions. Each question is worth 5 marks.

1. Show how the noisy channel model of Statistical Machine Translation (SMT) is derived from Bayes theorem.

2. Briefly explain (using an example) the following three uses of SMT: *1)* assimilation, *2)* communication and *3)* dissemination.

3. What is the difference between *fluency* and *adequacy*? Using English as the target language and a language of your choosing as the source language, provide an example of a translation that is *a)* fluent and adequate, *b)* fluent and inadequate, *c)* disfluent and adequate, and *d)* disfluent and inadequate.

4. Explain, using an example, the role of the *brevity penalty* in BLEU.

5. What is the main difference between the type of training data needed to train a translation model and the type of training data needed to train a language model? What role does the *Expectation Maximization* algorithm play in training a translation model?

6. Describe, using examples, two advantages that phrase-based SMT has over word-based SMT.

7. Define the notion of *future cost* in SMT decoding, and explain, using an example, why it might be useful to calculate this during decoding.

8. Briefly describe the *Word Error Rate (WER)* evaluation metric and provide an example of how it can unfairly penalize a good translation.

9. What do you understand by the term *linguistic knowledge?* Provide an example of how linguistic knowledge of the source or target language can benefit *a)* the translation process, and *b)* the evaluation process.

10. State two different ways two language models could differ from one another. What is the metric used to measure how effective a language model is at predicting a piece of text?

[End of Question 1]

QUESTION 2 ***[TOTAL MARKS: 30]***

Q 2(a) **[12 Marks]**

Given the following pairs:

Les chats *Des chats*

The cats *cats*

State what the following translation probabilities will be after *two* iterations of the Expectation Maximisation algorithm and show all the steps followed to arrive at these values:

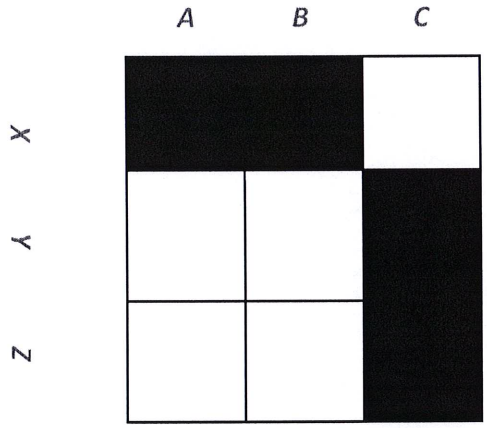
i. $t(cats|chats)$
ii. $t(cats|Les)$
iii. $t(cats|Des)$
iv. $t(The|chats)$
v. $t(The|Les)$
vi. $t(The|Des)$

Q 2(b) **[8 Marks]**

Given a source sentence, s, of length 4 and a target sentence, t, of length 2, how many possible alignments exist from t to s, assuming the existence of the NULL token? Justify your answer.

Q 2(c) **[10 Marks]**

List all phrase pairs that are consistent with the following word alignment:



[End of Question 2]

QUESTION 3 ***[TOTAL MARKS: 30]***

Q 3(a) **[10 Marks]**

State the equation for the probability of a sentence and show how this is approximated using *a)* a 4-gram language model, *b)* a 5-gram language model and *c)* a 7-gram language model. Provide one example from the English language which demonstrates how a trigram language model is superior to a bigram language model.

Q 3(b) **[10 Marks]**

Discuss the importance of *smoothing* in language modelling. List three smoothing methods that are based on the notion of *count adjustment*. Provide an example which highlights an inadequacy of using count adjustment alone to perform smoothing and briefly explain what other information can be taken into account during the smoothing process.

Q 3(c) **[10 Marks]**

Consider the following data:

Count	Count of Counts
1	8000
2	3000
3	1200
4	700
5	400

Readjust the following three bigram counts using *Good-Turing* smoothing.

Count	Bigram
4	Coffee break
3	Coffee cup
1	Coffee morning

[End of Question 3]

QUESTION 4 ***[TOTAL MARKS: 30]***

Q 4(a) **[12 Marks]**

Using diagrams, step through an example of how *stack decoding* works in conjunction with *histogram pruning*.

Q 4(b) **[10 Marks]**

Calculate the unigram, bigram, trigram and 4-gram precision, and the brevity penalty for the following translation:

Translation: The man on the truck came to the scene.
Reference: The man with the van arrived there.

Q 4(c) **[8 Marks]**

List the three types of evaluation used in MT research. Discuss the advantages and disadvantages of each type.

[End of Question 4]

[END OF EXAM]