

**A0 XXXX**

**THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING**

No calculator permitted in this examination

**THE UNIVERSITY OF BIRMINGHAM**

Degree of BSc with Honours  
Artificial Intelligence and Computer Science. Final Examination  
Computer Science/Software Engineering. Final Examination  
Computer Science/Software Engineering with Business Studies. Final Examination

Joint Degree of BSc with Honours  
Psychology and Artificial Intelligence. Final Examination

MSc in Computer Science

**06 11223**

Natural Language Processing and Applications

May/June 2002 2 hours

[Answer THREE Questions out of Five]

Turn Over

- 1 -

THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING

No calculator

**NOTE** Percentage marks given on the paper are rounded and so may not appear to add up to 100%.

*This question was given to you in advance.*

1. (a) “Machine translation is essential if Internet users around the globe are to converse on an equal footing.” (Modified from *The Computer Bulletin*, July 2000.)
- (b) “Speech is not just the future of Windows: it is the future of computing itself.” (Bill Gates)

Choose ONE of these quotations and discuss it. Your answer should focus on whether you agree with the quotation or not (and why) and on how far the implied technology is currently feasible.

[33%]

THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING

No calculator

A table of IPA symbols and phoneme features is given on Page 7.

2. (a) “Two of the allophones of the /t/ phoneme in English are the unaspirated [t] found in *still* [stɪl] and the aspirated [t<sup>h</sup>] found in *till* [t<sup>h</sup>ɪl].”

Explain carefully what is meant by the terms ‘allophone’ and ‘phoneme’ in this sentence, and hence what the sentence as a whole means.

[9%]

- (b) In some dialects of English, two of the phonological rules relating /t/ with its allophones can be written as:

/t/ → [t<sup>h</sup>] : word-boundary .. vowel  
/t/ → [d] : vowel .. vowel

It can be assumed that elsewhere the rule /t/ → [t] applies.

Briefly explain the notation used in stating these rules, and give (in IPA notation) the expected phonetic pronunciation of the words *Tate* /teɪt/, *rating* /reɪtɪŋ/ and *straight* /streɪt/.

[5%]

- (c) A particular speech synthesizer uses a two stage process: graphemes → phonemes, phonemes → allophones.

(i) Explain in detail how phonological rules of the kind given in (b) above can be used in the second stage of the speech synthesizer, sketching the algorithm(s) involved.

[7%]

(ii) Suggest TWO advantages and TWO disadvantages of this approach to speech synthesis, compared to direct conversion of graphemes to sounds by looking up stored phonetic representations.

[5%]

- (d) Speech recognition systems could work by the reverse process to that described in (c), i.e. recognizing sounds as allophones, and then generating written text via allophones → phonemes, phonemes → graphemes. Most commercial systems do not work in this way. How do they work? Why?

[7%]

THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING

No calculator

A table of IPA symbols and phoneme features is given on Page 7.

3. (a) Consider these sentences:

*We **agree** sometimes.*  
*He **agrees** with me.*  
*I wasn't **agreeing** with you.*  
*No-one **agreed**.*  
*Your **agreement** was required.*  
***Agreements** should be honoured.*

Using the emboldened words in the above sentences as examples, explain what is meant by the terms 'lexeme', 'morpheme', 'inflectional morphology' and 'derivational morphology'. Why is the distinction between 'inflectional' and 'derivational' relevant to computer processing of English morphology?

[16%]

- (b) The regular past tense is formed in written English by the addition of the letters *ed* to end of the 'base form' of the verb. This is sometimes combined with other changes, e.g. doubling the last letter of the base form or changing a terminal *y* to an *i*, as shown in the table below.

Base Verb	Past
<i>wait</i>	<i>waited</i>
<i>knit</i>	<i>knitted</i>
<i>cry</i>	<i>cried</i>
<i>play</i>	<i>played</i>

Outline the rules which are operating on the examples shown in the table, and then explain how these can be combined with an algorithm to generate the past form of a verb from the base form. Take into account irregular past forms (e.g. *eat – ate*).

[9%]

- (c) Consider the pronunciation of the words in the table below.

Word	Pronunciation
<i>climb</i>	[klaɪm]
<i>climber</i>	[klaɪmə]
<i>timber</i>	[tɪmbə]
<i>clamber</i>	[klæmbə]

(Note that *\*timb* and *\*clamb* are not modern English words.) Suggest reasons for the pronunciation of the letters *mb* in the words in the table. What implications does this have for speech synthesis?

[8%]

THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING

No calculator

4. The following fragment of a grammar describes some ‘simple sentences’ in a dialect of English spoken in the 16th century. (Note that *nom*, *acc*, *first* and *second* are constants; *P1* and *P2* are variables.)

$S \rightarrow \text{pro}(P1, \text{nom}) \text{ VP}(P1)$

$\text{VP}(P1) \rightarrow \text{verb}(P1) \mid \text{verb}(P1) \text{ pro}(P2, \text{acc})$

$\text{pro}(\text{second}, \text{nom}) \rightarrow \textit{thou}$

$\text{pro}(\text{second}, \text{acc}) \rightarrow \textit{thee}$

$\text{pro}(\text{third}, \text{nom}) \rightarrow \textit{he}$

$\text{pro}(\text{third}, \text{acc}) \rightarrow \textit{him}$

$\text{verb}(\text{second}) \rightarrow \textit{likest}$

$\text{verb}(\text{third}) \rightarrow \textit{liketh}$

$\text{verb}(\text{second}) \rightarrow \textit{walkest}$

$\text{verb}(\text{third}) \rightarrow \textit{walketh}$

- (a) For each sentence below, state whether it is valid or not according to the grammar given above. Using your answers as appropriate, explain in some detail how the grammar enforces agreement in person (second or third) and case (*nom* or *acc*).
- (i) *thou likest him*
  - (ii) *thou liketh him*
  - (iii) *he walketh*
  - (iv) *he liketh thou* [9%]

- (b) The grammar allows sentences like *thou walkest him* (modern English *you walk him*). Alter appropriate productions in the grammar above so that such sentences are not allowed (while still valid when the verb is *like*). [4%]

- (c) Explain how the grammar above can be extended to include a representation of the syntax tree. You do NOT need to re-write all of the productions in the grammar, but should give examples to show how this would be done. [7%]

- (d) Briefly explain how a machine translation system based on syntax trees works. The mistranslations given below are typical of those produced by such systems. Explain why they occur and suggest how they might be avoided.

French input: *Une femme a porté un pantalon vert*. English output: *A woman carried green trousers*. (The verb *porter* in French can mean either ‘carry’ or ‘wear’.)

English input: *One woman wore green trousers*. French output: *Une femme a porté des pantalons verts*. (The French literally means that one woman wore several pairs of green trousers; the singular *pantalon* means ‘a pair of trousers’.)

[13%]

THIS IS NOT AN OPEN-BOOK  
EXAMINATION – CANDIDATES MAY  
NOT CONSULT ANY REFERENCE  
MATERIAL DURING THE SITTING

No calculator

5. (a) Consider the first noun phrase in these sentences:

<i>Everest is a mountain.</i>	<i>*The Everest is a mountain.</i>
<i>*Thames is a river.</i>	<i>The Thames is a river.</i>
<i>*Atlantic is an ocean.</i>	<i>The Atlantic is an ocean.</i>
<i>Windermere is a lake.</i>	<i>*The Windermere is a lake.</i>
<i>London is a city.</i>	<i>*The London is a city.</i>
<i>England is a country.</i>	<i>*The England is a country.</i>

\* marks sentences which have incorrect uses of the determiner *the*. How might semantic features be used to ensure that incorrect uses of *the* are rejected? Suggest possible semantic features for the proper nouns above which will achieve this.

[7%]

- (b) Explain what is meant by the term ‘theta role’ ( $\theta$ -role), distinguishing carefully between participant and circumstantial theta roles.

[8%]

- (c) A proposed lexicon entry for the verb *give*, employing  $\theta$ -roles, is of the form:

*give* : complements =

NP/Agent: [ +ANIMATE ] + NP/Patient + PP (*to*)/Recipient [ +ANIMATE ] |  
NP/Agent: [ +ANIMATE ] + NP/Recipient [ +ANIMATE ] + NP/Patient

Explain what this entry means, showing how it may be used to produce semantic analyses of the following sentences:

*I gave a book to John.*  
*I gave John a cat.*

[9%]

- (d) Only the first of the following sentences is acceptable when spoken in a ‘normal’ intonation (e.g. without extra stress).

*I gave John a book, because it was his birthday.*  
*\*I gave a book to John, because it was his birthday.*

One explanation for this is that the ordering of the complements to the verb *give* carries presuppositions. Explain what these are. How might this be helpful in computer processing of the semantics of a text?

[9%]

<i>IPA</i>	<i>SEE Examples</i>	<i>ASCII</i>	<i>Partial Feature Set</i>
[i]	heel, me	IY	{vowel,voiced}
[ɪ]	hit	IH	{vowel,voiced}
[ɛ]	met, head	EH	{vowel,voiced}
[æ]	hat	AE	{vowel,voiced}
[ə]	about, after, fern	AX	{vowel,voiced}
[ʌ]	up, fun	UX	{vowel,voiced}
[u]	soon	UW	{vowel,voiced}
[ʊ]	put, foot	UH	{vowel,voiced}
[ɔ]	fork, taut	AO	{vowel,voiced}
[ɒ]	hot	OH	{vowel,voiced}
[ɑ]	bath, bar	AH	{vowel,voiced}
[eɪ]	wait, cake	EI	{vowel,voiced}
[aɪ]	kite, buy	AY	{vowel,voiced}
[ɔɪ]	coin, toy	OY	{vowel,voiced}
[oʊ]	bone, open	OU	{vowel,voiced}
[aʊ]	cow, out	AW	{vowel,voiced}
[ɪə]	ear, sheer	IA	{vowel,voiced}
[ɛə]	air, share	EA	{vowel,voiced}
[ʊə]	tour	UA	{vowel,voiced}
[p]	spin	p	{stop,bilabial,voiceless}
[b]	boo	b	{stop,bilabial,voiced}
[t]	stop	t	{stop,alveolar,voiceless}
[d]	dog	d	{stop,alveolar,voiced}
[k]	scan	k	{stop,velar,voiceless}
[g]	gate	g	{stop,velar,voiced}
[m]	mat	m	{nasal,bilabial,voiced}
[n]	not	n	{nasal,alveolar,voiced}
[ŋ]	king	NG	{nasal,velar,voiced}
[f]	fat	f	{fricative,labiodental,voiceless}
[v]	vat	v	{fricative,labiodental,voiced}
[θ]	thumb	TH	{fricative,dental,voiceless}
[ð]	that	DH	{fricative,dental,voiced}
[s]	sat	s	{fricative,alveolar,voiceless}
[z]	zip	z	{fricative,alveolar,voiced}
[ʃ]	mesh	SH	{fricative,palatal,voiceless}
[ʒ]	measure	ZH	{fricative,palatal,voiced}
[h]	hot	h	{fricative,glottal}
[tʃ]	chair	CH	{affricative,palatal,voiceless}
[dʒ]	edge, jam	JH	{affricative,palatal,voiced}
[l]	lot	l	{approximant,voiced}
[r]	rot	r	{approximant,voiced}
[j]	yawn	y	{approximant,voiced}
[w]	win	w	{approximant,voiced}

### Notes

bilabial = both lips

dental = tongue tip and upper teeth

palatal = tongue and hard palate

labio-dental = upper teeth and lower lip

alveolar = tongue tip and tooth ridge

velar = tongue and soft palate