

- 2 -

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

No calculator

NOTE All questions carry equal marks. 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. Answer ONE of the following questions:

- (a) Microsoft is reported to be integrating a “natural user interface” (NUI) into the next version of its Windows operating system. Compared to the now standard ‘graphical user interface’ (GUI), what natural language features would an ideal NUI have? How feasible is it to provide these features at present?
- (b) In 1998, a press release from CompuServe stated that machine translation of e-mails would be standard “in the near future”. Why has this not happened yet? What are the prospects for the next five years?

[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 11.

2. (a) “Two of the allophones of the /p/ phoneme in English are the unaspirated [p] found in *spit* [spɪt] and the aspirated [p^h] found in *pit* [p^hɪt]. By contrast, in French only the unaspirated [p] is found, whereas in Hindi the unaspirated [p] and the aspirated [p^h] correspond to different phonemes.”
- (i) Explain carefully what is meant by the terms ‘allophone’ and ‘phoneme’ in the above sentences, and hence what the sentences as a whole mean. [8%]
- (ii) Briefly explain the operation of a text-to-speech (TTS) system based on a two stage approach: graphemes to phonemes followed by phonemes to allophones (as used for example in the MacinTalk synthesiser). [7%]
- (iii) What are the consequences for TTS in English, French and Hindi of the information given above concerning the /p/ phoneme? [4%]
- (iv) What are the consequences for speech-to-text (STT) in English, French and Hindi of the information given above concerning the /p/ phoneme? (You do **not** need to explain how a speech recognition system might work.) [3%]

/ Question continues on the next page

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

No calculator

- (b) Consider the phonetic information given below for four pairs of words in a particular dialect of German. ([ç] is a phone not found in English: a soft velar fricative.)

<i>Word</i>	<i>Approximate Pronunciation</i>	<i>Word</i>	<i>Approximate Pronunciation</i>	<i>Approximate Meaning</i>
Bund	[bunt]	Bunde	[bundə]	bundle / bundles
bunt	[bunt]	bunte	[buntə]	colourful (singular / plural)
Tag	[tak]	Tage	[təgə]	day / days
gelb	[gɛlp]	gelblich	[gɛlbliç]	yellow / yellowish

- (i) Consider the pronunciation of the terminal stop in the first of each of the four pairs of words and how it changes in the second word in each pair. Construct a possible phonological rule based on this evidence, generalized to all stops. Explain the rule in your own words and write it out in feature set notation. [3%]
- (ii) Suggest how a German whose dialect included this rule might pronounce the English words *tig* (correct English pronunciation [tɪg]) and *tigging* (correct English pronunciation [tɪgɪŋ]) if he or she incorrectly applied the rule to English. Some native German speakers pronounce the English word *have* as [hæf] instead of [hæv]. Suggest a reason. [4%]
- (iii) Suppose it is desired to generate from English text speech which sounds to a native English speaker as if it were spoken by a German (e.g. for use in film dubbing). Explain how this might be done. [5%]

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 11.

3. (a) Consider these sentences.

*You must **ferment** grape juice to make wine.*
*Wine is made by **fermenting** grape juice.*
*Wine is made by **fermentation**.*
***Fermentations** proceed slowly in the cold.*
*There was quite a **ferment** when the change was announced.*

Note that the verb *ferment* is pronounced [fə: 'mɛnt] (i.e. with the stress on the second syllable), whereas the noun *ferment* is pronounced [fə: mɛnt] (i.e. with the stress on the first syllable).

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

[17%]

- (b) The regular past tense is formed in spoken English by the addition of the phones [d], [t] or [ɪd], as shown in the following examples.

Base Verb	Past
<i>tap</i> [tæp]	<i>tapped</i> [tæpt]
<i>tab</i> [tæb]	<i>tabbed</i> [tæbd]
<i>tuck</i> [tʌk]	<i>tucked</i> [tʌkt]
<i>tug</i> [tʌg]	<i>tugged</i> [tʌgd]
<i>wait</i> [weɪt]	<i>waited</i> [weɪtɪd]
<i>wade</i> [weɪd]	<i>waded</i> [weɪdɪd]

Assume that these examples are typical of all English verbs ending in stops and with regular past tense forms. Outline the rule(s) which are operating here, and then explain how they 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%]

/ Question continues on the next page

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

No calculator

- (c) Some consonant letter pairs in English have varying pronunciations, a fact which causes obvious difficulties for English text-to-speech (TTS) systems.
- The *g* in *gn* is sometimes silent (e.g. in *sign, signing, signed*) and sometimes pronounced (e.g. in *signal, signature*).
 - The *b* in *mb* is similarly sometimes silent (e.g. in *climb, climbing, climbed, climber*) and sometimes pronounced (e.g. in *clamber, timber*).
 - In SEE, the letters *ng* sometimes correspond to two phones [ŋ dʒ] (e.g. in *danger, plunger*) and sometimes to one phone [ŋ] (e.g. in *ring, ringer, ringing, sing, singer*).

The examples demonstrate the lack of independence between at least the sound and the morphological levels of natural language processing. Explain in some detail why this is so and what the consequences are.

[7%]

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 a noun phrase (NP) in French. (Note that m , f , s and p are constants; G and N are variables, standing for Gender and Number respectively; the underscore $_$ matches any constant or variable.)

$NP(G, N) \rightarrow \text{det}(G, N) \text{ SNP}(G, N)$

$SNP(G, N) \rightarrow \text{noun}(G, N) \text{ adj}(G, N)$

$SNP(G, N) \rightarrow \text{noun}(G, N)$

$\text{det}(m, s) \rightarrow \textit{le}$

(*le, la and les all mean the*)

$\text{det}(f, s) \rightarrow \textit{la}$

$\text{det}(_, p) \rightarrow \textit{les}$

$\text{adj}(m, s) \rightarrow \textit{amical}$

(*amical, amicaux, amicale and amicales all mean friendly*)

$\text{adj}(m, p) \rightarrow \textit{amicaux}$

$\text{adj}(f, s) \rightarrow \textit{amicale}$

$\text{adj}(f, p) \rightarrow \textit{amicales}$

$\text{noun}(m, s) \rightarrow \textit{bébé}$

(*bébé means baby*)

$\text{noun}(m, p) \rightarrow \textit{bébés}$

(*bébés means babies*)

$\text{noun}(f, s) \rightarrow \textit{femme}$

(*femme means woman*)

$\text{noun}(f, p) \rightarrow \textit{femmes}$

(*femmes means women*)

- (a) For each noun phrase below, state whether it is valid or not according to the grammar given above. Using the noun phrases as examples where appropriate, explain in some detail how the grammar enforces agreement in gender (masculine or feminine) and number (singular or plural) in both generation and recognition.
- (i) *le bébé amical*
 - (ii) *les bébés amicales*
 - (iii) *les femmes amicale*
 - (iv) *les femmes amicaux* [8%]
- (b) 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. Write down or draw the syntax tree for the French noun phrase *la femme amicale*. [7%]

/ Question continues on the next page

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

No calculator

- (c) (i) The equivalent grammar for an English noun phrase lacks the enforcement of agreement in gender. Write down the grammar (excluding syntax trees) which will generate correct English noun phrases from the words *the*, *friendly*, *baby*, *babies*, *woman* and *women*. Write down or draw the resulting syntax tree for the noun phrase *the friendly woman*. [4%]
- (ii) Explain how grammars and trees of the kind considered above can be used to translate between English and French noun phrases, e.g. between *la femme amicale* and *the friendly woman*. Make sure your answer includes a discussion of gender and number. [8%]
- (iii) A particular machine translation system translates *One woman wore the green trousers* into the French sentence *Une femme a porté les pantalons verts*. However, *pantalon* is singular in French, so that the French translation literally means that she wore more than one pair of green trousers. What does this suggest about the method of machine translation employed by the system? How might it be corrected? [7%]

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

No calculator

5. (a) In translating weather forecasts, it has been suggested that the words *rain*, *snow*, *hail* and *sleet* should all be marked by the semantic feature +PRECIPITATION.

- (i) Using the above information as an example, explain what is meant by the term 'semantic feature'.
- (ii) How could this and similar semantic features be used to attempt to resolve the referent of the word *it* in translating the following two pairs of sentences taken from weather forecasts?

London will have snow today. It will fall to a depth of 10 cm.
London will have rain today. It will continue tomorrow.

[4%]

- (b) Explain briefly what is meant by the term 'theta role' (θ -role). Illustrate the distinction between syntactic roles and theta roles by considering the noun phrase *the player* in the sentences:

The player stopped the ball.
The ball was stopped by the player.

[7%]

- (c) A proposed lexical entry for the verb *stop* (in the syntax used in this module) is:

stop : complements =
NP/Agent: [+ANIMATE] + NP/Patient: [-ANIMATE] + {PP (*with*) /Instrument}

Use this lexical entry to explain IN DETAIL how a computer program could produce semantic analyses of the sentences:

- (i) *The umpire stopped the game with his whistle.*
(ii) *The game was stopped by the umpire.*
(iii) *The game was stopped.*

Your answer should include consideration of the passivization rule in English.

[13%]

/ Question continues on the next page

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

No calculator

- (d) The following two sentences suggest that the lexical entry for *stop* is not complete.
- (i) *The rain stopped the game.* With a syntactic subject like *rain*, no Instrument is possible.
 - (ii) *The game stopped.* There is no syntactic object (i.e. no noun phrase after the verb).

Suggest additional lexical entries and/or changes to your answers to (c) above and explain briefly how these can produce appropriate analyses of the two sentences and equivalent passive forms where these exist.

[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, fem	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,alveolar,voiceless}
[ʒ]	measure	ZH	{fricative,alveolar,voiced}
[h]	hot	h	{fricative,glottal}
[tʃ]	chair	CH	{affricative,alveolar,voiceless}
[dʒ]	edge, jam	JH	{affricative,alveolar,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