UNIVERSITI SAINS MALAYSIA

First Semester Examination
Academic Session 2002/2003

September 2002

CCS503 – Intelligent Document Processing

CSC514 – Natural Language Processing

Duration : 3 hours

INSTRUCTION TO CANDIDATE:

- Please ensure that this examination paper contains **FOUR** questions in **SIX** printed pages before you start the examination.
- Answer **ALL** questions.
- This is an Open Book Examination.
- You can choose to answer either in Bahasa Malaysia or English.

ENGLISH VERSION OF THE QUESTION PAPER

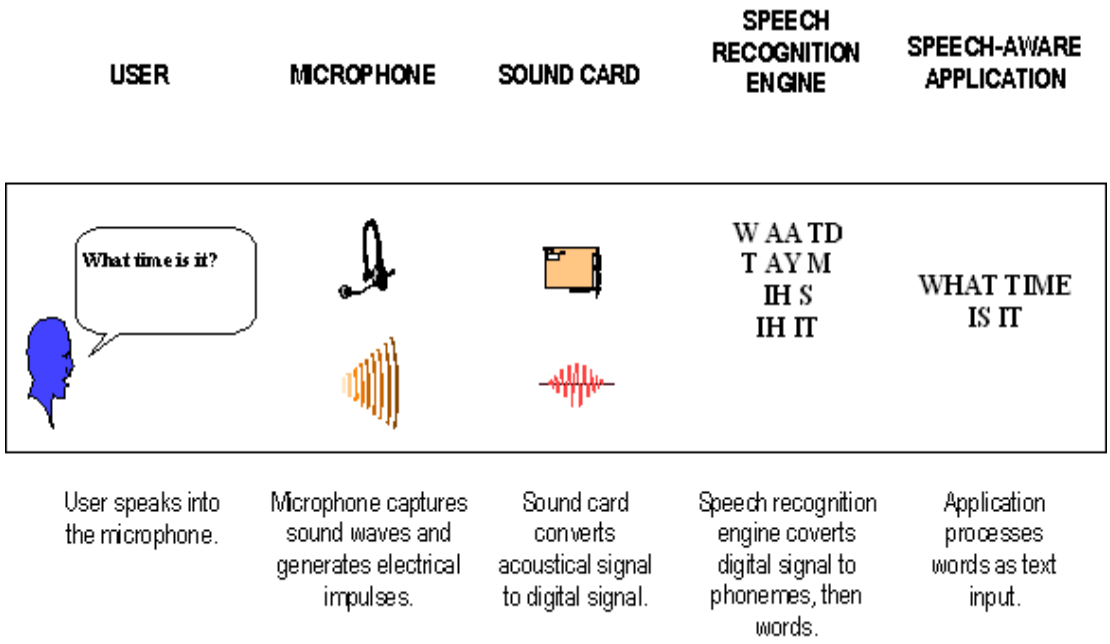
1. Given the following sentence:

"What time is it?"

(a) How many phonemes are there in the given sentence? Write down the pronunciation of the sentence given above in IPA symbols.

[25/100]

(b) Speech recognition, or speech-to-text, involves capturing and digitizing the sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analyzing the words to ensure correct spelling for words that sound alike (such as write and right). The figure below illustrates this high-level description of the process.



Discuss in detail how the various NLP tools introduced in this course can be applied to improve the performance of the speech recognition engine as mentioned above.

[50/100]

(c) In English, the regular way of forming the plural of a noun is to add an 's' to the end of the word. Such 's' is pronounced differently (i.e. /z/ or /s/) in different contexts. For example, the plural of *bid* is [bɪdz] and the plural of *bit* is [bɪts]. Construct the appropriate phonological rule(s) for this phenomenon.

[25/100]

2. This question has two parts. Both parts MUST be answered.

(a) A noun may be expressed in the singular or plural form, e.g. *cat ~ cats*, *house ~ houses*. In English, plurality is expressed by various inflections.

baby ~ babies	*maze ~ mazes*
box ~ boxes	*rash ~ rashes*
bush ~ bushes	*shelf ~ shelves*
buzz ~ buzzes	*spy ~ spies*
cat ~ cats	*switch ~ switches*
chicken ~ chickens	*tax ~ taxes*
echo ~ echoes	*toad ~ toads*
frog ~ frogs	*tomato ~ tomatoes*
glass ~ glasses	*watch ~ watches*
house ~ houses	*wife ~ wives*
knife ~ knives	*wolf ~ wolves*

(i) Using ONLY the data provided above, determine the various possible inflections in English.

[5/100]

(ii) The inflections are used only under certain conditions. Now write rules to generate the appropriate plural form in English. Your rules should be in the following format. Remember to give the condition(s) under which the rule may apply.

singular_ending $\rightarrow$ plural_ending | condition(s)

[40/100]

(iii) Indicate the order in which the rules are to occur.

[5/100]

(b) In Spanish, nouns too may be expressed in the singular or plural form, e.g. *libro ~ libros*, *iglú ~ iglúes*. As with English, plurality is expressed by various inflections. [In Spanish, accent on a syllable is indicated by *′*, as in *el bambú* 'bamboo', *el ladrón* 'robber' and *el lápiz* 'pencil'.]

el bambú	'bamboo'	*los bambúes*	'bamboos'
el corredor	'runner'	*los corredores*	'runners'
el hombre	'man'	*los hombres*	'men'
el iglú	'igloo'	*los iglúes*	'igloos'
el joven	'young man'	*los jóvenes*	'young men'
el ladrón	'robber'	*los ladrones*	'robbers'
el lápiz	'pencil'	*los lápices*	'pencils'
el libro	'book'	*los libros*	'books'
el lunes	'Monday'	*los lunes*	'Mondays'
el mujer	'woman'	*las mujeres*	'women'
el nivel	'level'	*los niveles*	'levels'
el padre	'parent'	*los padres*	'parents'
el papel	'paper'	*los papeles*	'papers'
el paraguas	'umbrella'	*los paraguas*	'umbrellas'
el pollo	'chicken'	*los pollos*	'chickens'
el régimen	'regime'	*los regímenes*	'regimes'
el rubí	'ruby'	*los rubíes*	'rubies'
la bici	'bike'	*las bicis*	'bikes'
la casa	'house'	*las casas*	'houses'
la ciudad	'city'	*las ciudades*	'cities'
la flor	'flower'	*las flores*	'flowers'
la nación	'nation'	*las naciones*	'nations'
la pluma	'pen'	*las plumas*	'pens'
la universidad	'university'	*las universidades*	'universities'
una vez	'one time'	*algunas veces*	'a few times'

(i) Using ONLY the data provided above, determine the various possible inflections in Spanish.

[5/100]

(ii) The inflections are used only under certain conditions. Now write rules to generate the appropriate plural form in Spanish. Your rules should be in the following format. Remember to give the condition(s) under which the rule may apply.

singular_ending → plural_ending | condition(s)

[40/100]

(iii) Indicate the order in which the rules are to occur.

[5/100]

3. Given the following grammar:

S → noun(SEM0) VP(SEM0)
VP(SEM0) → verb(Comps, SEM0, SEM1) VerbComps(Comps,SEM1)
VerbComps(none,SEM1) → Ø
VerbComps(noun,SEM1) → noun(SEM1)
noun(SEM0) → {noun, SEM0}
noun(SEM1) → {noun, SEM1}
verb(Comps, SEM0, SEM1) → {verb, Comps, SEM0, SEM1}

lexicon:
cows : noun, +ANIMATE
grass : noun, -ANIMATE
eat : verb, noun, +ANIMATE, -ANIMATE
sleeps : verb, none, +ANIMATE, -

(a) List all sentences generated by the above grammar which are semantically valid but grammatically incorrect (in English).

[20/100]

(b) Extend the grammar and lexicon given above so that only grammatically correct (in English) sentences will be generated.

[30/100]

(c) List all sentences generated by the grammar given in 3(b) which are grammatically correct and semantically valid (in English).

[15/100]

(d) Construct a chart produced by the bottom-up chart parsing technique for one of the longest sentences listed in 3(c) above. Values of all variables involved are required to be shown in the chart.

[35/100]

4. (a) For each of the following NLP tools, describe its functionality and give an example input/output pair.

(i) Spelling checker

(ii) Inflectional morphological analyzer

(iii) Part of speech tagger

(iv) Parser

(v) Word sense tagger

[40/100]

(b) Discuss in detail how the NLP tools in 4(a) can be applied to each of the following NLP applications.

(i) Search engine

(ii) Grammar checker

(iii) Intelligent helpdesk system (or call center)

[60/100]

- oooOooo -