UKLO Round 2 March 2012

Data and questions

[Use the answer sheet for all questions.]

1. A fox among the h (15 marks)

Spencer is working on an online biology reference, and he is currently working on the information retrieval program, so that people can type in things like "What do whales eat?" or "How much does a bee weigh?" and get the relevant answers.

Part of this task involves a process called stemming – taking text and figuring out what the "stem" of each word is. (The stem is the form of the word without any prefixes or suffixes, so dance is the stem of dancing, happy is the sum of unhappiness, etc.) The program needs the stem so that it can determine that a request about "walruses" needs data from the article WALRUS, and that one about "fungi" needs data from the article FUNGUS).

So Spencer writes a series of rules for determining the singular form of plural nouns. He writes a rule "Remove final S" to handle cats → cat, a rule "Replace final I with US" to handle fungi → fungus, a rule "Remove final E" to handle vertebrae → vertebra, and so on. He ends up with the following rules:

Remove final S	Remove final EN
Replace final ICE with OUSE	Replace final A with UM
Replace IES with Y	Replace final I with US
Remove final E	

When he applies his program to a series of real words, however, it doesn't always work. Here are some outputs of his program:

INPUT	INTENDED OUTPUT	ACTUAL OUTPUT
cats	cat	cat
dogs	dog	dog
walruses	walrus	walrus
foxes	fox	fox
oxen	ox	ox
bacteria	bacterium	bacterium
fungi	fungus	fungus
horses	horse	hors
chimpanzees	chimpanzee	chimpanze
algae	alga	algum
guppies	guppy	guppi
hens	hen	h
mice	mouse	mous

Q1.1. What output would Spencer's program produce for the following words?

a. bees
b. kiwis
c. flies
d. fleece
e. geese

Q1.2. What went wrong with the program?

Q1.3. What can you determine about the order in which Spencer's program applied the rules?

Q1.4. Could putting the rules in a different order cause the program to work? If so, what is the order?

2. Who is good? (15 marks)

Luiseño is a Uto-Aztecan language spoken in southern California. Although it has only about 30 or 40 native speakers now, there is an active campaign to teach it to children. It has no standard written form, and is not widely used in writing except in schools. One way to represent it is to use the International Phonetic Alphabet, as in the following examples; by convention, IPA symbols are enclosed in square brackets. The pronunciations don't matter for the problem, but you may like to know that:

[ʔ] is a glottal stop, the sound in the middle of *uh-oh*.
[q] is like a k but made further back in the mouth.
[ʃ] is a "sh" sound as in *shout*
[j] is a "y" sound as in *yellow*.
[ʂ] is like an "s", but made with the tongue-tip curled back.
[x] is a "ch" sound roughly as in Irish *deich*, or Scots *loch*, or German *Bach*.
[ŋ] is the nasal "ng" sound in *singer*.
[:] indicates a long vowel.

This transcription shows how words are pronounced, but when we speak we don't normally pause between words, so it is traditional not to show word-breaks in a phonetic transcription.

Here are some sentences in Luiseño and their English translations.

1	[nawitmalqajwukalaqpoki:k]	The girl does not walk home.
2	[jaʔaʃpolo:v]	The man is good.
3	[hu:ʔunikatqajtʃipomkat]	The teacher is not a liar.
4	[haxʂuxetʃiqʂuŋa:li]	Who hits the woman?
5	[jaʔaʃwukalaq]	The man walks.
6	[to:wqʂuʂuŋa:lihu:ʔunikat]	Does the teacher see the woman?
7	[ʔiviʂuŋa:lnona:jixetʃiq]	This woman hits my father.
8	[nona:jiʂuxetʃiqʔiviʂuŋa:l]	Does this woman hit my father?
9	[ʔiviʂuŋa:lxetʃiqnona:ji]	This woman hits my father.
10	[hu:ʔunikattʃipomkat]	The teacher is a liar.
11	[ʔivihu:ʔunikatnona:jito:wq]	This teacher sees my father.
12	[hu:ʔunikatʂuto:wqʂuŋa:li]	Does the teacher see the woman?

Q2.1. Translate the following into English.

a. [jaʔaʃwukalaqpoki:k]

b. [xetʃiqʂuʂuŋa:lnona:ji]

c. [haxʂuqajtʃipomkat]

d. [ʂuŋa:liʂuto:wqhu:ʔunikat]

Q2.2. Translate the following into Luiseño.

e. Is the teacher a liar?

f. The teacher sees the woman.

g. This girl does not see my father.

h. Who is good?

3. The little engine that could… read (15 marks)

Professor Monotone's "Astounding Linguistic Knowledge Engine for Making Inferences" (ALKEMI) can, when given a list of true statements, deduce further true statements from it. For example, if it knows that "Professor Monotone can read Russian", it can deduce that "Professor Monotone can read". We represent this as:

> Professor Monotone can read Russian $\Rightarrow$ Professor Monotone can read

This means that whenever the first statement is true, the second has to be true, too; there's no way for the first to be true while the second is false. We call this a legitimate inference.

The Professor's machine can go through statements and, by making particular sorts of changes, generate further statements that follow from them. However, it's not as easy as replacing "can read Russian" with "can read" anywhere you find it. For example, funny things happen when the statement contains one of a set of words called "quantifiers", including *every*, *some*, *no*, *a*, *few*, *many*, *three*, and so on.

> No student can read Russian $\Rightarrow$ No student can read   WRONG!

The inference is not legitimate: even if no student can read Russian, it is entirely possible that students can read Japanese, English or Spanish.

Each of the quantifiers allows a different pattern of legitimate inferences, so the professor's machine keeps a special table of patterns and uses it to derive new statements from given ones. We've reproduced it below. It may look mysterious, but given the information in this table and a list of inferences produced by the machine shown on the next page, you can work out what each part means and how the machine works.

	Quantifier	Side	Direction
A	Every	Left	Downward
B	Every	Right	Upward
C	No	Left	Upward
D	No	Right	Downward
E	Some	Left	Upward
F	Some	Right	Upward

Unfortunately, however, there is one error in the table above that is causing the professor's machine to make some illegitimate inferences!

Here are some examples of inferences declared legitimate by the professor's machine:

- Every teacher can read English $\Rightarrow$ Every English teacher can read English
- Some English students can read English $\Rightarrow$ Some English students can read
- No English student can read Russian $\Rightarrow$ No student can read Russian
- Every teacher can read English and Russian $\Rightarrow$ Every teacher can read Russian
- No student can read Russian $\Rightarrow$ No student can read English and Russian
- Every teacher can read English $\Rightarrow$ Every Russian teacher can read English
- Some Russian students can read English $\Rightarrow$ Some students can read English
- No English student can read $\Rightarrow$ No English student can read English

Q3.1. Which table row (A-F) contains a mistake and caused the machine to draw one or more illegitimate inferences? (2 points)

Q3.2. The list of inferences isn't complete. The professor's machine could draw additional inferences as well. Using only words that appear in the table of quantifiers above, generate another legitimate inference that the machine could have drawn from "Every teacher can read English". (4 points)
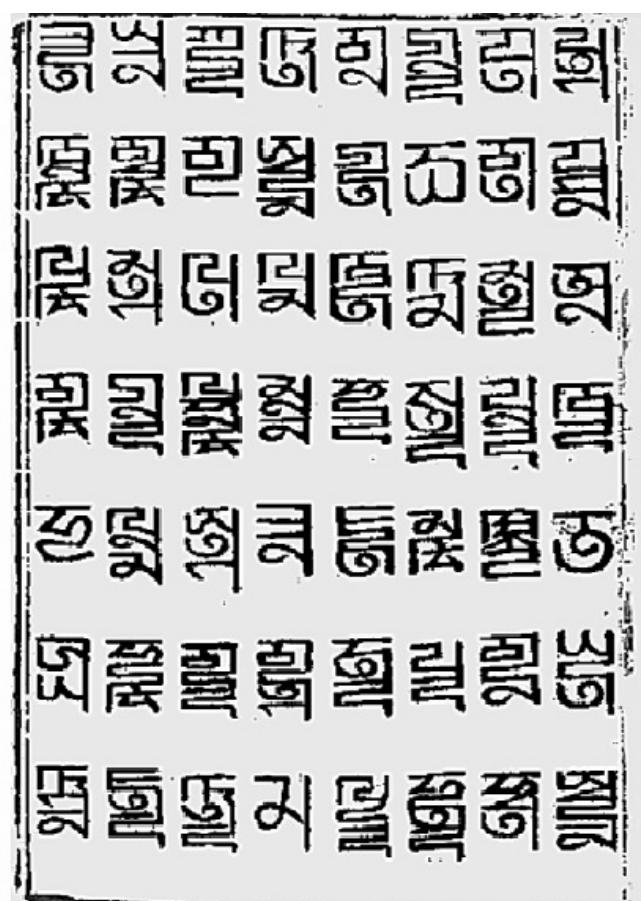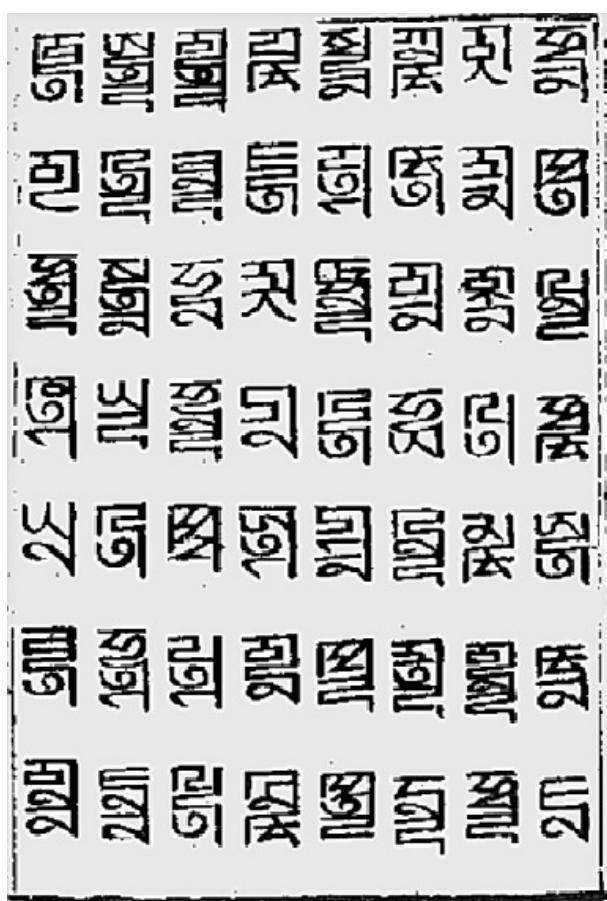
Q3.3. The professor's machine doesn't yet understand every quantifier. Help it learn the quantifiers *at least three, at most three*, and *not all* by completing the table below in your answer book: put "Upward" or "Downward" in the appropriate cells. (9 points)

	Quantifier	Side	Direction
G	At least three	Left	
H	At least three	Right	
I	At most three	Left	
J	At most three	Right	
K	Not all	Left	
L	Not all	Right	

4. 100 surnames [25 marks]

When the Mongol Emperor Kublai Khan initiated the Yuan dynasty (1271–1368) in China, he commissioned Lama ´Gro-mgon Chos-rgyal ´Phags-paa to create a unified script to write all the major languages under his rule. Although the resulting system (now called ´Phags-pa) never caught on beyond official use, some classic Chinese texts survive in a ´Phags-pa version.

The Bǎijiāxìng (Hundred Surnames) is a Song Dynasty (960–1279) poem listing over 400 classical Chinese family names. Although originally written in Chinese characters, during the Yuan dynasty this poem was written in ´Phags-pa characters as well, as shown below.

The illustration shows two consecutive pages of the Bǎijiāxìng Měnggǔwén ("The Hundred Surnames in Mongol Script"), from a 1340 manuscript.

Below are twenty lines (9-28) from the Yuan-era Bǎijiāxìng, with some names missing (identified by row number and column letter). The two pages given above correspond to a portion of the poem below. Your task is to figure out which portion of this poem the pages represent, and use this to figure out what the missing names must be.

|	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
9	Fi	Lem	Drxim	Sĭa,	Lue	Ho	Yi	Thang
10	Dxing	'In	Lo	Pi,	Haŭ	'U	'An	Srang
11	Yaŭ	Yĭu	Sri	Fu,	Bue	Pen	Dzi	Khang
12	U	Yĭu	Ngĭŭan	Pu,	Ku	Mung	Bing	Hŏang
13	**[13a]**	Fu	Sring	Taĭ,	Dam	Sung	**[13g]**	Bang
14	Xĭung	Ki	**[14c]**	Khĭu,	**[14e]**	Trĭu	Tung	Lĭang
15	Du	**[15b]**	Lam	**[15d]**,	Zi	**[15f]**	**[15g]**	Gĭang
16	Kĭa	Lu	Lxiŭ	Ngue,	Kĭang	Dung	**[16g]**	Kŭaŭ
17	**[17a]**	Sring	Lim	Xĭaŭ,	Trung	Zĭu	Khiŭ	Laŭ
18	Kaŭ	**[18b]**	Tshaĭ	Den,	Fan	Hu	**[18g]**	Faŭ
19	Ngĭu	Wan	Tri	Ko,	**[19e]**	Kŏan	Lu	Maŭ
20	Kĭing	**[20b]**	**[20c]**	Wu,	Kan	Xĭaĭ	'Ing	Tsung
21	Ting	Sĭŭan	Pue	Dxing,	'Ĭu	Sren	Hang	Hung
22	Paŭ	Trĭu	**[22c]**	Sri,	Tshue	Kĭi	Nriŭ	Kĭung
23	Dring	Xĭi	Xĭing	**[23d]**,	Bue	Lĭu	Ngĭung	'Ung
24	Sĭun	Yang	**[24c]**	Xĭue,	Trin	Khĭu	Kĭa	Fung
25	Nyue	Yi	Drĭu	Kin,	Ki	Ping	Mue	Zĭung
26	Tsing	Dŏan	Fuŭ	Wu,	'U	Tsĭaŭ	Pa	Kĭung
27	Wu	Ngue	Sran	Ku,	Trhĭa	Hiŭ	Fu	Bung
28	Dzĭŭan	Trhi	Pan	Ngĭang,	Tshiŭ	Drĭung	Yi	Kĭung
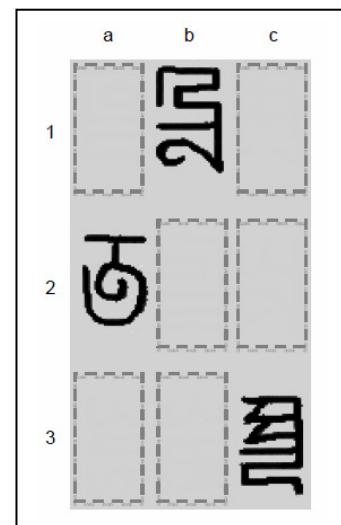
Values correspond to Yuan dynasty pronunciation rather than modern pronunciation:

- r indicates the previous consonant is pronounced with a curled tongue tip.
- h indicates that the previous sound is pronounced with an extra puff of air.
- ny is as in 'onion', ng as in 'finger'.
- ' is a glottal stop, the sound in the middle of 'uh-oh'.
- x indicates a sound like the 'h' in 'huge'.
- A vowel with a ˘ mark means that it is a "glide" – a short vowel-like sound transitioning into or out of the syllable's main vowel: so ŭ and ŏ are like a 'w', while ĭ is like a 'y'.

Q4.1. In your answer book, fill in the 18 missing names. (10 points)

Q4.2. On the right is a partial 3x3 excerpt from one larger page of a 1418 manuscript of the Băijiāxìng Měnggŭwén. Six of the names have been left out. In your answer book, draw them in the spaces provided. (9 points)

Q4.3. Explain how the ´Phags-pa writing system works. (6 points)

5. Catalan plurals (25 marks)

Catalan is the official language of Catalonia, in north-east Spain. The plural of Catalan nouns is usually formed by adding the ending -s. But if the noun ends in one of the letters s, x or ç, more complex rules apply. Here are the singular and the plural forms of some Catalan nouns (in simplified spelling) and their English translations. Some forms are missing.

singular	plural	trans-lation	singular	plural	translation
el apèndix	els apèndixs	appendix	el ònix	**[f]**	onyx
el bastaix	**[a]**	carrier	el pàncrees	els pàncrees	pancreas
el troleibús	**[b]**	trolley-bus	el pedaç	els pedaços	patch
el cactus	els cactus	cactus	la pelvis	**[g]**	pelvis
la càries	les càries	caries	el permís	els permisos	permission
[c]	les clos	meadow	el pis	**[h]**	flat, lodgings
el contumaç	els contumaços	rebel	el **[i]**	els sequaços	(male) follower
la faç	les faços	face	la **[j]**	les sequaçs	(female) follower
el flux	els fluxos	stream	el sufix	els sufixos	suffix
el gimnàs	els gimnasos	gym	**[k]**	els tastaolletes	frivolous person
la hèlix	les hèlixs	screw	el teix	els teixos	yew
el índex	els índexs	index	la trencadís	les trencadís	piece of crockery
el iris	**[d]**	rainbow	el vas	els vasos	vase
el llaç	**[e]**	loop	la xeix	**[l]**	(the letter) x

Q5.1. Fill the gaps. Explain your solution.

Q5.2. The marks ` and ′ indicate the stress in certain Catalan words. Find out when they are used.

Q5.3. Which syllable is stressed in those words where there is no ` or ′ mark?