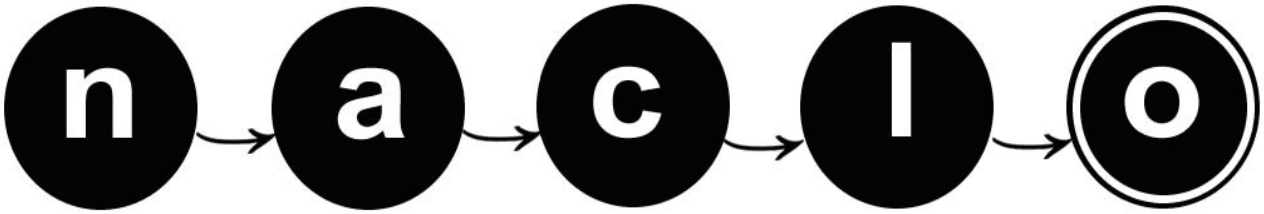


The Second Annual North American Computational Linguistics Olympiad

Solutions

for Invitational Round

Problem F: Willie Costello
Problem G: Partick Littell
Problems H: Eric Breck
Problem I: Patrick Littell
Problem J: Todor Tchervenkov
Problem K: Patrick Littell and Erin Donnelly
Problem L: Adam Hesterberg
Booklet editor: Eugene Fink

(F) Fakepapershelfmaker

F1. The following is a list of several Japanese words with their English meanings; use them to write definitions of the Japanese compounds.

sakura	cherry blossom	***kami***	paper	***nise***	fake
shiru	soup	***tana***	shelf	***tsukuri***	maker
iro	color(ed)	***tanuki***	raccoon	***hako***	box

(a) ***nisetanukijiru***	fake soup made out of raccoons
(b) ***nisedanukijiru***	soup made out of fake raccoons
(c) ***irogamibako***	box for colored paper
(d) ***irokamibako***	colored box for paper
(e) ***nisezakuradana***	shelf for fake cherry blossoms
(f) ***nisesakuradana***	fake shelf for cherry blossoms

F2. Match the following four-member Japanese compound words with their English meanings; one of the Japanese words has two possible meanings.

(1) a fake shelf-maker made of paper	B: ***nisekamitanadzukuri***
(2) a maker of fake shelves for paper	D: ***nisekamidanadzukuri***
(3) a fake maker of shelves for paper	D: ***nisekamidanadzukuri***
(4) a shelf-maker made of fake paper	C: ***nisegamitanadzukuri***
(5) a maker of shelves for fake paper	A: ***nisegamidanadzukuri***

F3. Explain your answers.

When we compound two Japanese words, the first word modifies/describes the second. For example, adding *hashi* before *hako* makes a word meaning a box (*hako*) for chopsticks (*hashi*). As another example, adding *nuri* before *hashi* makes a word meaning chopsticks (*hashi*) that are lacquered (*nuri*).

Every simple (noncompound) word has two forms: the basic form, used when it occurs alone, and the variant form, sometimes used in compound words.

Basic	**Variant**
hako	*bako*
hana	*bana*
hashi	*bashi*
kami	*gami*
kiri	*giri*
sakura	*zakura*

Basic	**Variant**
shiru	*jiru*
sora	*zora*
tana	*dana*
tanuki	*danuki*
tsukuri	*dzukuri*

The variant form has a different first letter, which depends on the first letter in the basic form. Specifically, we replace the initial *h* with *b*, initial *k* with *g*, initial *s* with *z*, initial *sh* with *j*, initial *t* with *d*, and initial *ts* with *dz*. As a side note, some letters do not require replacement, but they do not occur in the problem.

We next deduce rules for compounding simple words; we denote basic forms by *a*, *b*, *c*, and *d*, and respective variants by *a*, *b*, *c*, and *d*. We first notice that two-member compounds have the following structure:

$$a + b \rightarrow a\underline{b}$$

Three-member compounds have two different structures, which depend on their meaning. If we first form a word containing *a* and *b*, and then compound it with *c*, we use the following structure:

$$(a + b) + c \rightarrow a\underline{b} + c \rightarrow a\underline{bc}$$

If we first compound *b* and *c*, and then add *c*, we use a different structure:

$$a + (b + c) \rightarrow a + b\underline{c} \rightarrow ab\underline{c}$$

Thus, when we combine two (simple or compound) words into a larger compound word, we use the following rules:

- We use the original form of the first word.
- If the second word is simple (noncompound), we use its variant form.
- If the second word is compound, we do not change it.

When compounding four simple words, we can get five different internal structures; two of them give the same result, which is why the four compounds in Problem F2 correspond to five possible meanings.

We can now determine which English version corresponds to what structure.

(1) **a fake shelf-maker made of paper**
→ fake + (paper + (shelf + maker))
→ a + (b + (c + d))
→ a + (b + c*d*)
→ a + b*cd*
→ ab*cd*
→ ***nise-kami-tana-dzukiri*** (B)

(2) **a maker of fake shelves for paper**
→ (fake + (paper + shelf)) + maker
→ (a + (b + c)) + d
→ (a + b*c*) + d
→ ab*c* + d
→ ab*cd*
→ ***nise-kami-dana-dzukuri*** (D)

(3) **a fake maker of shelves for paper**
→ fake + ((paper + shelf) + maker)
→ a + ((b + c) + d)
→ a + (b*c* + d)
→ a + b*cd*
→ ab*cd*
→ ***nise-kami-dana-dzukuri*** (D)

(4) **a shelf-maker made of fake paper**
→ (a + b) + (c + d)
→ a*b* + c*d*
→ ab*cd*
→ ***nise-gami-tana-dzukuri*** (C)

(5) **a maker of shelves for fake paper**
→ ((a + b) + c) + d
→ (a*b* + c) + d
→ (ab*c*) + d
→ ab*cd*
→ ***nise-gami-dana-dzukuri*** (A)

(G) Manam, I'm Anam

G1. Onkau's, Mombwa's, and Kulu's houses have already been located on the map. Who lives in the other five houses?

A: ***Pita*** B: ***Butokang*** C: ***Sulung*** D: ***Tola*** E: ***Sala***

G2. Arongo is building a house in the location marked with an X. In three Manam Pile sentences, describe this location in relation to the three closest houses.

1. ***Arongo pera kana ilau ieno, Butokang pera kana auta ieno.***
2. ***Arongo pera kana ata ieno, Pita pera kana awa ieno.***
3. ***Arongo pera kana awa ilau ieno, Sulung pera kana ata auta ieno.***

G3. Explain your answers.

The ananlysis of the given examples suggests that *auta, ilau, ata,* and *awa* are the significant words, which probably represent directions. For reference, "*X pera kana*" means "X's house", and *ieno* means "is located."

We can see that *auta* and *ilau* appear to be opposed, and that *ata* and *awa* are also opposed. We thus hypothesize that they represent two axes of dimenstions, and we support this hypothesis by observing that their compounds are intermediate directions, such as *awa ilau* vs. *ata auta*, and *awa auta* vs. *ata ilau.* In fact, these compounds may occur in either order; for example, *ilau awa* and *auta ata* are also directions. *Ilau awa* is similar but not identical to *awa ilau*, in the same way as "north-north-west" is similar but not identical to "west-north-west."

When we analyze the relative locations of the houses of Onkau, Kulu, and Mombwa, we may be tempted to assume that *auta* is North, *ilau* is South, *awa* is East, and *ata* is West. This assumption works until about halfway through the problem, but then we should notice contradictions: either these directions are very imprecise or some houses are in the sea.

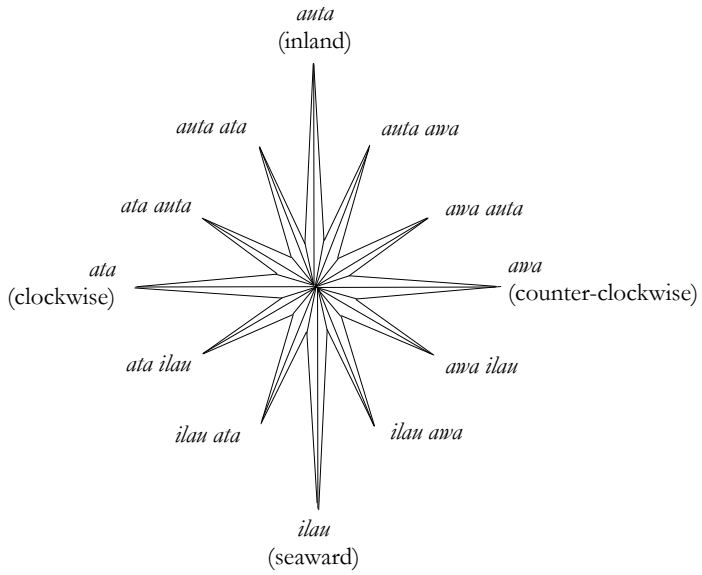
When we reach a contradiction, we should try discarding some of the underlying assumptions; in this case, we discard the assumption that the islanders reckon the traditional directions, that is, North, South, East, and West. Instead, we should consider other directional possibilities that may occur to the islanders.

In fact, *auta* means "inland" or "upland," which is the same thing on a cone-shaped volcanic island, and *ilau* means "seaward." Furthermore, *Ata* means "clockwise around the island," and *awa* means "counterclockwise". The compound direction *awa auta* thus means "inland in a counterclockwise direction".

An alternative approach to solving this problem is as follows. We may be fairly certain that the directions form two axes, *auta/ilau* and *ata/awa*. Instead of placing islanders on the given map, as soon as we have a hunch where they live, we can work out an abstract two-dimensional map indicating the relative locations of the houses. Then, by comparing it to the given map, we can see that the only way to reconcile the

two maps is to "wrap" the abstract map around the island, that is, to curve the Cartesian grid of houses into a polar grid centered on the volcano.

The full Manam compass rose is as follows:



Note that some of the directions are irrelevant to the problem, and we have included them only for completeness. Also note that the angle between *auta* and North depends on a specific location, which means that this compass would rotate with respect to the traditional North/South compass as we walk around the island.

If you have solved this difficult problem, you are probably able to examine and revise your initial assumptions, which is an essential research skill.

(H) Thorny Stems

1.	If a word ends in ***ies***, then replace ***ies*** with ***y***.	Exception: **series**
2.	If a word ends in ***ss***, then replace ***ss*** with ***ss***.	No exceptions
3.	If a word ends in ***ives***, then replace ***ves*** with ***fe***.	Exception: **hives**
4.	If a word ends in ***ves***, then replace ***ves*** with ***f***.	Exception: **caves**
5.	If a word ends in ***oes***, then replace ***oes*** with ***o***.	Exception: **floes**
6.	If a word ends in ***s***, then replace ***s*** with __.	Exception: **guesses**
7.	If a word ends in ***ing***, then replace ***ing*** with __.	Exception: **closing**
8.	If a word ends in ***ied***, then replace ***ied*** with ***y***.	Exception: **lied**
9.	If a word ends in ***ed***, then replace ***ed*** with __.	Exception: **posed**
10.	Otherwise the word is its own stem.	Exception: **formulae**

The order of rules is somewhat flexible, and the only requirements are as follows:

- Rules 1–5 are before Rule 6.
- Rule 3 is before Rule 4.
- Rule 8 is before Rule 9.

We may find multiple exceptions to most rules; some examples are as follows:

- Rule 5: toes
- Rule 6: bus
- Rule 7: ring
- Rule 9: bed

Notes and common mistakes

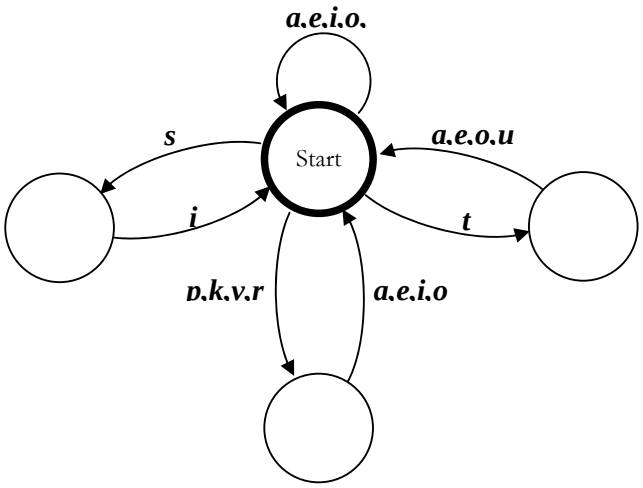
- The problem statement does not ask to include exception for the last ("otherwise") rule, and thus the grading has not accounted for this exception.
- The word "wives" has a unique pattern, which requires its own rule (Rule 3).
- We need Rule 2 so that Rule 6 does not remove ***s*** from words like "moss."
- Some contestants have listed more specific cases after more general cases, such as Rule 8 after Rule 9. Since we use the first matching rule, this ordering leads to ignoring the appropriate specific rule.
- An additional rule for the words ending in ***es*** is not required in the given rule set, since these words match either Rules 3–5 or Rule 6.
- A word is an exception only if the entire rule set gives a wrong result for this word. For example, "knives" is not an exception to Rule 4, because it matches Rule 3, which is before Rule 4.
- The problem statement does not allow the use of wildcards or other complex specifications; for example, we cannot collapse Rules 1 and 8 to a single ***ie**** rule, and we also cannot define a single rule for the words ending in **<consonant>*ves***.
- Some contestants have indicated that a word like "princess" may be stemmed to "prince", but it is a different sort of word change.
- Several contestants have solved a different problem; specifically, they have given a list of rules and exceptions that together cover the given word list, and thus they have included words from the given list as exceptions. While this problem is also interesting, we issued a clarification that the competition problem was different.

(I) aw-TOM-uh-tuh

I1. Identify possible Rotokas words.

Possible: ***iu, oire, urioo, raorao, uaia***
Impossible: ***idau, uente, voav, oratreopaveiepa***

I2. Specify path labels so that exactly half of the words below succeed.



oisio	*tiravau*	*saiuu*	*kotoe*
uasau	*usta*	*sioparoia*	*parauos*
puapuata	*sisigarue*	*porouativeve*	*aasiia*

I3. Why do *t* and *s* get their own edges? What is special about these letters?

The letters *t* and *s* have restrictions on their use before vowels, which makes them different from the other consonants. We can use *s* only before *i*; for example, "sisigarue" is a valid word, whereas "uasau" is invalid. Furthermore, we cannot use *t* before *i*, which means that we can use it only before *a*, *e*, *o*, and *u*; for example, "kotoe" is a valid word, whereas "tiravau" is invalid.

This observation implies a special relationship between *t* an *s*; in fact, it suggests that these two sounds are the same on an abstract level, although their pronounciation and spelling depends on the following vowel.

(J) The Curragh of Kildare

J1. Determine the Irish names of the following villages and translate each name.

	English	Irish	Translation
20	Mullaghbane	***An Mullach Bán***	***The White Summit***
21	Killananny	***Cill an Eanaigh/ Coill an Eanaigh***	***Church of the Fen/ Wood of the Fen***
22	Knocknakillardy	***Cnoc na Cille Airde/ Cnoc na Coille Airde***	***Hill of the High Church/ Hill of the High Wood***
23	Gortnabinna	***Gort na Binne***	***Field of the Peak***
24	Clashgortmore	***Clais an Ghoirt Mhóir***	***Pit of the Big Field***
25	Killbeg	***An Chill Bheag/ An Choill Bheag***	***The Small Church/ The Small Wood***
26	Blackcastle	***An Caisleán Dubh***	Black castle

J2. Explain your answers.

Orthographic correspondences: The English names are phonetic imitations of the Irish names. The letter correspondences (Irish/English) include *c/k*, *ch/gh*, and *aigh/y*, but many Irish letters do not have English equiavalents; for example, there is no distinction between *cill* and *coill*.

Irish place names: The names fit the following pattern, where brackets represent optional parts; note that adjectives come after the respective nouns:

[An] <noun-1> [<adjective-1>] [an/na <noun-2> [<adjective-2>]]

If a name includes a second noun, it is in the "of" form, which is analogous to the "<noun>'s" form in English, such as "John's." If it includes an adjective after the "of" noun, this adjective is also in the "of" form. Furthermore, an article before the "of" noun is sometimes *na* rather than *an*. We can identify the related patterns by comparing the two forms.

Nouns:

Base form	"Of" form	Trans-lation
gort	an ghoirt	field
an currach	an churraigh	marsh
an pháirc/páirc	na páirce	park
cill	na cille	church
an choill	na coille	wood
an bun/ bun	?	base
an bhinn	?	peak
baile	?	town
cluain	?	meadow
gleann	?	valley
eanach	?	fen

Base form	"Of" form	Trans-lation
an dún	?	ford
talamh	?	land
an mhainistir	?	abbey
an chlais	?	pit
?	na muice	pig
?	an mhullaigh	summit
?	an uain	lamb
?	an chairn	mound
?	an chaisleáin	castle
?	an chnoic	hill

We notice two classes of nouns.

Class A: The nouns whose last vowel is *i*.
- Insert *–h–* in the base form when preceded by the article.
- Add *–e* in the end to construct the "of" form.
- Use the article *na* in the "of" form.

Class B: The nouns whose last vowel is not *i*.
- No changes in the base form.
- Add *–i–* before the last consonant cluster to construct the "of" form.
- Use the article *an* and insert *–h–* after the first consonant in the "of" form.

Adjectives: The behavior of an adjective depends on the class of the related noun.

Base form		**"Of" form**		**Translation**
Class A	Class B	Class A	Class B	
	dhubh		duibhe	black
	bhán	bháin		white
ard				high
	íseal			low
mór	mhór			big
beag				small

An adjective after a Class A noun behaves like a Class A noun with an article. Similarly, an adjective after a Class B noun behaves like a Class B noun with an article.

English place names: The Irish words always have the same English correspondence, regardless of their grammatical form, with the exception of the *–ach/–aigh* words; for example, *bán*, *bháin*, *bhán*, and *báine* all correspond to *–bane* in an English name.

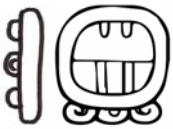
(K) Tzolk'in

K1. Draw the Mayan names of the days labeled *a* and *b* on the calendar (see next page).

a. 

b. 

K2. Write *c* and *d* on the calendar for the following days (see next page).

c. 

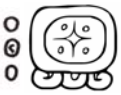
d. 

K3. How often does the following day occur?

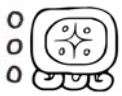
e. 

Every 260 days.

August

						
						
				***c***		
						

September

	b					
						
			***d***			
					*a*	

K4. Explain your answers.

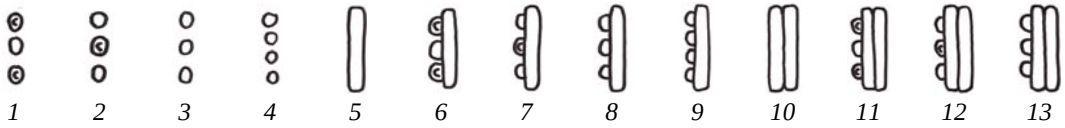
We first observe that each day name includes two glyphs, which repeat in cycles of different length. In particular, the right-hand side glyphs, which look like little pictures, repeat every twenty days. For example, the  glyph repearts three times: August 18, September 7, and September 27; as a side note, it means Venus. Thus, the picture glyphs should repeat either every 20 days, or in some shorter cycle, which is a divisor of 20; however, if we consider all smaller divisors of 20, we find out that they cause "collisions" between glyphs, which means that the length of the cycle is exactly 20. As another side note, there is no way to identify the beginning of this cycle, and Mayans do not have a general consensus about its "start" day.

On the other hand, the glyphs on the left appear to cycle every 13 days:

We find three missing glyphs in Problems K2 and K3, and we can use the observed pattern to put them in their proper places. We can also deduce the positions of days in Problems K2c and K3d, which increases the certainty of placing the dot-and-bar glyphs.

We can determine the "start" for this sequence by observing the pattern of these glyphs; specficially, the arrow above shows the discontinuity in the pattern, which is likely to be the start of the cycle, thus leading to the following order:

We next observe that (1) the third glyph consists of three empty circles, and (2) the eighth glyph has three cirlces, the ninth has four circles, and the tenth has an extra bar instead of the circles. In fact, these symbols are numbers, and we can deduce their representation; specifically, a number is a sum of its elements, where an empty circle is 1 and a bar is 5. We can thus deduce that the fourth glyph consists of four empty circles:



We can now determine most day names.

In particular, September 28 (Problem K1a) is as shown on the left, since it is immediately after a "3" day, and it is 40 days after August 19, which should have the same picture glyph.

The only places for glyphs in Problems K2c and K2d are on August 23 and September 19, which gives the positions of 19 out of the 20 picture glyphs, and leaves only one missing picture glyph.

The day in Problem K1b falls on one of the missing-glyph days. We next note that the missing picture glyph appear in Problem K3, and thus it should be as shown on the left.

In conclusion, we observe that the lengths of the two cycles, 20 and 13, are relatively prime, which means that the length of the combined cycle is $20 \cdot 13 = 260$ days, and thus the Tzolk'in year is 260-day long. Note that it serves only as the ritual calendar, and not as the agricultural calendar.

Notes

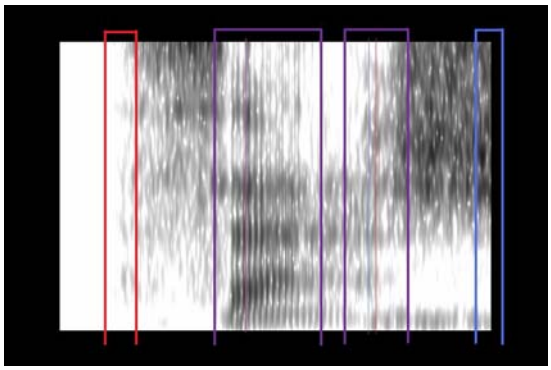
Some contestants have noticed that the picture glyphs 9 and 11 days after "Venus" are identical. This use of two identical glyphs is a typo; the two glyps are similar, and the problem editor has accidentally used the same symbol. Fortunately, it does not lead to any critical contraditction, and it does not make the problem unsolvable. We apologize for this typo, but we also wish to notice that linguists sometimes encounter similar problems in their research, since the authors of ancient records also made mistakes.

(L) The Whole Spectrum

L1. What words are shown in the last four spectrograms?

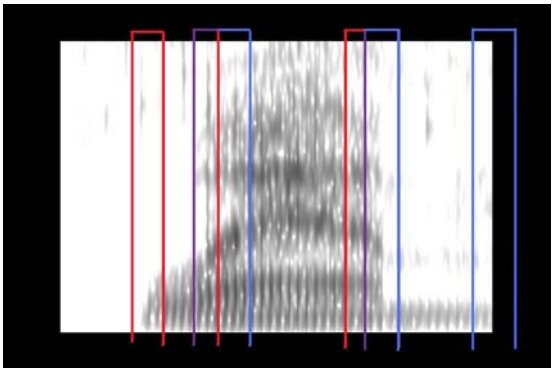
13: Lease 15: Sheep
14: Ash 16: Louse

L2: Mark the intervals corresponding to the three most significant sounds in "sash."



Each left endpoint is correct if it is between the corresponding pair of thick red lines, and each right endpoint is correct if it is between the corresponding pair of thick blue lines. Ideally, there should be significant overlap between the intervals affected by /*s*/ and /*a*/.

L3: Do the same for lamb.



As above. The intervals for /*l*/ and /*a*/ must overlap or be adjacent, and the intervals for /*a*/ and /*m*/ must overlap, since the /*m*/ nasalizes the end of the preceeding /*a*/.

L4: Explain your answers.

When the same basic sound, which is called *phoneme*, occurs in multiple words, it has similar effects. In particular, *s* and *sh* have distinctive appearances, and the vowels have distinctive sets of bars, which are called *formants*. These formants are effected by adjacent consonants in most cases (in fact, some consonants, called *stop consonants* (e.g. *p* and *k*) can be distinguished only this way, hence the seeming lack of a /*p*/ after "sheep"), which

suffices to identify the first three spectrograms. Note that the apparent shifting of the formants in the first one does not indicate a diphthong, but is simply a glide from /*i*/ to /*j*/. Also, the difference between /*i*/ (lease) and /*ai*/ (lice) is only an initial /*a*/, so the similarity between the end of, say, "ice" or "mice" and spectrogram 13 is not relevant. The vowel in the last one is not one shown in the previous spectrograms: just as /*ai*/ (as in "mice" or "shine") shifts from /*a*/ to /*i*/, the vowel in the last spectrogram shifts from /*a*/ to /*u*/. The English vowel with this property is /*au*/, so the last spectrogram is of "*louse*."

L5: Discuss the correspondence beteen the spellings and spectrograms of the given words.

Vowels clearly affect particularly long intervals, as do the sibilants *s* and *sh*, which can be said both continuously and loudly. Nasals after vowels also affect long intervals, because they nasalize the preceding vowels, although since the quality of some copies of the problem made this impossible to see, it was not graded. Transitions between sounds are not instantaneous, since the mouth changes smoothly from one position to another, so "intervals" have at least some level of imprecision. For most sounds in this problem, transitions were fairly abrupt, but others, e.g. final stops (initial stops are visible by their aspiration, an initial region of high amplitude) are detectable mostly by their effects on adjacent vowels, although they seem to have no intervals to themselves. With this observation in mind, it is possible to postulate another stop consonant at the start of vowel-initial words, and careful pronunciation of them does indeed reveal an unwritten glottal stop. In the given spectrograms, a glottal stop is present in every vowel-initial word except *e* (the beginning of *e* in the given spectrograms is simply a matter of amplitude, although credit was given for glottal stops even if the only cited example was *e*, as long as some description was given). Also, some diphthongs are arguably not indicated, and the glides that come after English long vowels, such that /*j*/ after /*i*/ and /*w*/ after /*u*/, are not indicated, but clearly visible at the end of, say, "knee." Conversely, certain letters of English orthography are not pronounced at all, such as final silent *e*, the initial *k* in "knee," the final *b* in "lamb," the doubled letters in words like "coo," and other vowel combinations. Most of these were pronounced at one point in the history of English, but as pronunciations changed, the orthography did not follow it.