

**North American Computational Linguistics
Olympiad (NACLO) 2007**

SOLUTIONS TO THE PROBLEMS

(DRAFT)

This is only a draft.

A. We are all molistic in a way

None of the adjectives are real English words. There are two classes of adjectives: "bad" and "good". We will refer to this property of adjectives as "polarity".

Each sentence links two or more adjectives as follows: "X and Y" indicates that X and Y have the same polarity. "X but Y" means that they have opposite polarities. Furthermore, "X and not Y" indicates opposite polarities, "even though X, Y" also indicates opposite polarities, while "not only X but also Y" associates adjectives of the same polarity.

The sentence about Diane shows that "strungy" and "struffy" are positive (desirable) quantities. By identifying other occurrences of the same words in other sentences, one can label each adjective as either positive or negative.

There are seven positive adjectives (strungy, struffy, cloovy, frumsy, danty, cluvious, and brastic) and five negative ones (weasy, blitty, sloshful, slatty, molistic).

A1. Only sentence c. includes adjectives of the right polarities, given the structure of the sentence.

A2. Only answer d. ("frumsy") is on the positive list above.

B. Pooh's Encyclopedia

The retrieval is based on simple keyword matching: the search engine compares the word roots in a given query with those in article titles, and identifies the titles that have at least two word roots in common with the query. Note that, when matching irregular verbs, it determines their roots based on their present tense (e.g. "write" for "wrote"). The matches for the questions by Pooh and his friends are as follows; the matching words are marked by capital letters.

Winnie-the-Pooh:

Query: Where should a BEAR STOCK his jars of honey?

Match: Lost tales of "Bulls vs. BEARS" STOCK trading

Query: How much honey should a BEAR store for the WINTER?

Match: WINTER hibernation of BEARS and rodents

Eeyore:

Query: Where should I LOOK for my LOST tail?

Match: Ways to LOOK for LOST things

Query: Which ANIMALS SLEEP during the winter?

Match: Effects of honey on the SLEEP quality of humans and ANIMALS

Christopher Robin:

Query: What is the shortest WAY from my place to the HOUSE of Winnie-the-Poor?

Match: WAYS to store food in the HOUSE

Query: Who wrote the BOOKS about Pooh BEAR?

Match: BOOKS about care and feeding of BEARS

C. A Donkey in every house

C1:

Align the sentences: (2 points)

Greek Sentence English Sentence

A	5
B	6
C	2
D	3
E	1
F	8
G	7
H	4

Explanation: (1 point)

In order to align the Ancient Greek sentences with the English sentences, you have figure out the content words (master, son, donkey, house, and slave) and the singulars and plurals. In order to get started, you need an anchor. Once you have an anchor, you can figure out the rest by logic and process of elimination.

Various anchors are possible. Three are described here.

1. Notice that four English sentences contain the word "master" or "masters" and that four Greek sentences contain words that start with "cyr". No other word occurs four times. Therefore, "master" would be "cyr".
2. Count singulars and plurals. For example, in five English sentences, the second noun is plural and five Greek sentences have the word "ton".
3. Although you can do this problem without recognizing any words, you might have recognized a few. For example "adelphoi" looks like "Philadelphia", the city of brotherly love. If you know that "phil" means "love" as in "bibliophile" (book lover), then you would know that "adelphoi" means brother. You might also notice that "emporoi" reminds you of the word "emporium", which is a market place.

C2

Translations (7 points)

("o:" is the vowel that is spelled as an "o" with a bar over it in the test booklet.)

the houses of the merchants

hoi to:n emporo:n oicoi

the donkeys of the slave

hoi tu dulu onoi

Explanation (5 points)

Vocabulary:

hyi son

dul slave

cyri master

oic house

on donkey

adelph brother

empor merchant

Order of words:

Each sentence starts with two articles, which are followed by two nouns. The first article starts with "h". The second article starts with "t". The first noun is the owner, and the second noun is the thing that is owned.

Number (singular and plural):

For the owner (first noun in Greek; second noun in English):
"o:n" is plural and "u" is singular.

For the owned (second noun in Greek; first noun in English):
"oi" is plural and "os" is singular.

Matching of articles and nouns:

The first article has an ending that matches the owned noun:

"ho" is singular and "hoi" is plural.

Examples:

ho ... dulos

the ... slave (singular)

hoi ... cyroi

the ... masters (plural)

The second article matches the owner:

"tu" is singular and "to:n" is plural.

Examples:

tu cyriu

the master (singular)

to:n hyio:n

the sons (plural)

Translations:

the houses of the merchants

hoi to:n emporo:n oicoi

Start with "hoi" because the owned noun (houses) is plural.

The next word is "to:n" because the owner (merchants) is plural.

The next word is the owner, which will be the root "empor" with the plural ending "o:n".

The next word is the owned noun, which will be the root "oic" with the plural ending "oi".

the donkeys of the slave

hoi tu dulu onoi

Start with "hoi" because the owned noun (donkeys) is plural.

The next word is "tu" because the owner (slave) is singular.

The next word is the owner, which will be the root "dul" with the singular ending "u".

The next word is the owned noun, which will be the root "on" with the plural ending "oi".

D. Hmong

U	a	V	c		
U	ai	P	h		
U	au	L	hl	°	g
U	aw		k		
U	ee	L	l		s
H	ev	P	m		
N	i	U	n		v
U	o	K	nts		
U	o	E	qh		
U	oo	H	r		
U	u	U	y		
U	ua				
J	w				

Solution: The syllables are written from right to left, as in the Roman script; however, within a syllable the vowel is written first, then the consonant. (This is so because Shong Lue Yang felt that the vowel was the more prominent sound, and the consonant a mere modification.) There is a letter for each vowels, and also letters for all consonants except **k**, which is pronounced by default if no consonant is indicated. The tone is indicated as a superscript mark above the vowel; both Shong Lue Yang's script and the missionaries' leave one of the four tones unmarked, but their choices are different.

(a)

9. $\overset{\circ}{U}U$ **noog**
10. $\overset{\circ}{J}V$ **cw**
11. $\overset{\circ}{U}P$ $\overset{\circ}{L}L$ **hais lus**
12. $\overset{\circ}{U}E$ $\overset{\circ}{U}P$ N **qhov muag kiv**

(b)

13. **hluav** $U L$
14. **li cas** $N L \overset{\circ}{U} V$
15. **neeg ntse** $\overset{\circ}{U} U \overset{\circ}{H} K$
16. **yawg** $\overset{\circ}{U} U$

E. Better sorry than shunk

She used to shink possums.

Now she shinks groundhogs for a living.

When she was in Eugene, she shank thirty-three possums in one day.

Then she took us possum-shinking in the Cascades.

This is the most likely set of forms for this verb, because of the relatively large number of real verbs that work this way in English, e.g., drink, drinks, drank, drinking, drunk; shrink, shrinks, shrank, shrinking, shrunk; sing, sings, sang, singing, sunk; sink, sings, sank, sinking, sunk, etc. These can serve as analogical models for new verb forms, e.g., children sometimes say things like "I brang my new toy" on this analogy.

E2: There are many, potentially an infinite number of, possible solutions to E1. The second most likely solutions are based on the analogy of other real verbs that have a "short u" sound in the form that follows "had", e.g.

shank, shanks, shunk, shanking, shunk based on hang, hangs, hung, hanging, hung (the alternate conjugations of this verb take "hanged" after "have," e.g., "They have already hanged the murderer.")

shink, shinks, shunk, shinking, shunk based on dig, digs, dug, digging, dug.

shunk, shunks, shank, shunking, shunk based on run, runs, ran, running, run. (This is less likely because there is only one verb in English that acts this way).

Much less likely:

shunk, shunks, shunk, shunking, shunk base on cut, cuts, cut, cutting, cut. (This is less likely because this class of real verbs in English all end in t or d, not k or g.

Even less likely, there may be any number of random forms of this verb, say yerkle, blumbles, jambolick, borging, shunk. Since this is a nonsense verb, and some verbs (like "to be" and "to go") are very irregular in English, it is impossible to limit the possible forms it could take. However, this solution is extremely unlikely, since in fact no verbs in English are totally random in their patterns, and those that are nearly so (like "to be" and "to go") are verbs that are used very often. Presumably "to shink/shunk" would not be such a common verb.

F. The lost tram

F1. The deviations in each text fragment are marked in bold and corrected:

(1)

The **tram (->train)** makes no stops; you sit **clown (->down)** and are served; there are no further intrusions, no **late-corners (->late-comers)**, no one hurrying to get off. The businessmen leaf through their financial reports, the lady with the hatbox is alone with her novel and her sirloin. Diners reading: you never see that on a plane. When the coast approaches **arid (->and)** dinner is over, everyone retires to his compartment to **he (->be)** transferred to the boat in peace, horizontally.

(Sunrise With Seamonsters, by Paul Theroux)

(2)

Usually, Howie could legitimately claim to have no **dear (->fear)** of any man or beast... Howie knew in his heart that it was **he (->the)** vulnerable positions he ended up in that scared him. He was used to operating from a position of strength, either real or projected. Now here he was, injured and alone, standing with **and (->an)** empty handgun in an open **filed (->field)**, while **hid (->his)** opponent or opponents **fried (->fired)** their weapon from behind solid cover.

(Rough Justice, by Mark Johnstone)

(3)

Two other factors **effect (->affect)** the body's temperature regulation: age and acclimatization. As we grow older, we **loose (->lose)** our ability to quickly regulate temperature... Very small children are also subject to heat disorders. **There (->Their)** small size allows them to take on heat much faster **then (->than)** adults. They also cannot indicate their thirst, **accept (->except)** through irritability. They are completely dependent upon adults to make certain they get enough fluids.

(Doctor in the House: Your Best Guide to Effective Medical Self-Care, by John Harbert)

F2. In the first text fragment, graphically similar letters or letter combinations are mixed up: **m<->in, m<->rn, ri<->n, cl<->d, h<->b**. This might have occurred if the text (probably messily printed or handwritten) had been interpreted by a computer (using an OCR, optical character recognition software) or (less probably) by a human who hadn't been paying attention to what he had been reading.

In the second text fragment, letters are skipped, added, rearranged or replaced by other letters (in the latter case, the pairs of letters corresponded to neighboring keys on a standard QWERTY keyboard: **d<->f, d<->s**). This most probably occurred when someone was typing too fast.

In the third text fragment, there are several lexical errors, when words with identical or very similar pronunciation are mixed up. This might have occurred if the person who had copied the text was quite bad in spelling, or maybe if the text was analyzed by a speech recognition system.

F3. Common spellchecking programs would not be of much help, since all wrong words are still English words (maybe the texts had already been through a spellcheck). To find at least a partial solution to fixing such deviations, one might create huge lists containing (1) common OCR mistakes (pairs of graphically similar words), (2) common misprints,

(3) commonly confused words. Some such lists already exist. Then, one could trace some (probably not all) mistakes using two alternative approaches. First, one could parse the texts using a natural language processing system, which might find some grammatical (mostly syntactical) mistakes. Constructing such systems is a very topical issue in modern computational linguistics, and a very complicated task. Second, one could verify all suspicious word combinations by searching them in a large text corpus, database, or simply in the web, and compare the number of hits to that of the alternative combination found in the lists. For example, a Google search yields some 8,690,000 results for **sit down**, and only 252 results for **sit clown** (probably most of them containing the same OCR error). This approach, however, only works for frequent word combinations and could accidentally result in wrong corrections for some rare, but not erroneous combinations. Therefore, the program should be an interactive one, marking potential mistakes and offering the user a variety of ways to correct them, but not attempting to correct them automatically.

G. Rewrite me badd

Proto-Tangkhulic form:	-ru (“bone”)	-khuk (“knee”)	-ko (“nine”)
Rule 1: K-Deletion			
Intermediate form 1:	-ru	-khu?	-ko
Rule 2: K-Insertion			
Intermediate form 2:	-ruk	-khu?	-ko
Rule 3: V-Raising			
Huishu form:	-ruk	-khu?	-ku

H. This problem is pretty//easy

Introduction: There are two things going on in the example sentences that are given in the problem statement. One is a change in meaning that is potentially disastrous:

1. You don't need to come // early.
2. Take the turkey out at five // to four.
3. I got canned // peaches.

The second is a confusion factor caused by a change in sentence structure:

4. All Americans need to buy a house // is a lot of money.
6. Fat people eat // accumulates in their bodies.

H1: Example sentences (5 points)

Your example sentences need to meet some minimal criteria:

1. The part before // should be a complete sentence.
2. The full sentence has a different meaning than the part before //.
- 2a. The part before // should not already be ambiguous.

H2: Ranking (4 points)

You were asked to rank two sentences that you made up along with sentences 4, 5, and 6.

4. All Americans need to buy a house // is a lot of money.
5. Melanie is pretty // busy.
6. Fat people eat // accumulates in their bodies.

If you take the confusion factor into account, 4 is the most confusing, followed by 6, and then 5.

H2: What makes a garden path sentence harder to process? (6 points)

All garden path sentences are either surprising or confusing, but what makes some harder than others? Looking at sentences 1-6, you might observe a number of things.

1. Change in part of speech: "fat" changes from an adjective in "fat people eat" to a noun in "fat accumulates in their bodies".
2. Change in structure: When you hear "fat people eat", you think that "eat" is the main verb of the sentence. When you hear "accumulates in their bodies", you realize that "people eat" modifies

"fat" and that the main verb of the sentence is "accumulates".

3. Missing words: 4 and 6 would become more clear if the word "that" were inserted:

All that Americans need to buy a house is a lot of money.
Fat that people eat accumulates in their bodies.

4. Intonation: 4 and 6 could be clarified with intonation.

5. Number of words before //: 6 has more words before // than 4 does.

6. Plausibility of the part before //: If you hear a complete and plausible sentence before //, you are less likely to expect more words. "All Americans need to buy a house" is a very plausible thing to say and is a complete sentence. "Fat people eat" is a generic statement, and you might want to hear more, so you might be expecting more words.

7. Words change meaning: "canned" can mean "fired" or "stored in a can".

8. Level of surprise: "I got canned" meaning "I was fired" could be a very surprising thing to say, and it is quite different from talking about groceries such as "canned peaches".