

(10 points)

(A) Gelda's House of Gelbelgarg (1/3)

A frequent problem in computational linguistics is that passages often use words that the computer simply doesn't have in its dictionary. Online slang evolves very fast, people use foreign words in English passages, people make typos and invent new abbreviations, etc. You could add new words to the dictionary as fast as you can find them and the next day the program could still be stumped by a new one!

But the program doesn't have to give up – instead, it can try to work out as much as it can. Various clues can tell a program whether something is a noun or a verb, a person or an inanimate object, etc., and you can even work out more! The following is a webpage where customers have rated their most recent experience at Gelda's House of Gelbelgarg. Even if you've never heard of any of these dishes, you can still figure out some things about them...

AI. Based on the following reviews, attempt to categorize the following items into:

I: Individual, discrete food items

L: Liquids, undifferentiated masses, or masses of uncountably small things

C: Containers or measurements

You won't be able to categorize them with 100% certainty, but use the category that you think is most probable for each. Choose a single category for each word below.

	I	L	C
färsel-försel			
gelbelgarg			
gorse-weebel			
rolse			
flebba			
göngerplose			
meembel			
sweet-bolger			



(A) Gelda's House of Gelbelgarg (3/3)

[xMandee7x](#)

Reviews: 4

I found the food confusing and disorienting. Where is this from? I randomly ordered the färsel-försel and had to send them back! Three words: weird, weird, and weird.

[Report this](#)

Food	★
Service	★★★
Atmosphere	★★★
Value	★

[wrldTrvl1977](#)

Reviews: 11

I went to Wolserl last year for a holiday, and this is the real thing. If you order the gelbelgarg, though, make sure you also get at least one rolse of sweet-bolger – it's how the locals like it!

[Report this](#)

Food	★★★
Service	★★
Atmosphere	★★★★
Value	★★★

[money@home](#)

Reviews: 103

the prices are steep, but i can afford them – i make up to \$75/hr working at home! find out how i do it at <http://bit.ly/grhCm>

User is on probation

Food	★★★
Service	★★★
Atmosphere	★★★
Value	★★★

[bu_zhidao](#)

Reviews: 8

not a great date spot! i got a gelbelgarg and a rolse of meembel, but my date was so disoriented that she just ended up with some gorse-weebel. :/

[Report this](#)

Food	★★
Service	★★
Atmosphere	★
Value	★★

[wembley2000](#)

Reviews: 2

The food was pretty good... But I would have liked more gorse-weebel and fewer göngerplose. You really feel like the chef is skimping on the good stuff..

[Report this](#)

Food	★★★
Service	★★
Atmosphere	★★★
Value	★



2010 Solutions

(A) Gelda's House of Gelbelgarg (I/I)

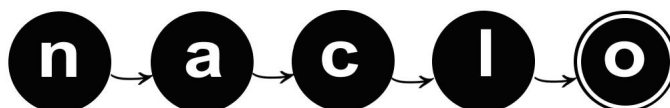
English systematically differentiates classes of nouns between whether they're Count – that is, are treated grammatically as if they can be counted, like *five cows* – or whether they're considered Mass, which can't themselves be counted. (This is a grammatical property of the *words*, not the items in question – even though rice comes in individual pieces you can't refer to five of them as “five rices” – you have to specify some measure word like “five *grains* of rice”.)

Mass nouns tend to be liquids, undifferentiated masses, or masses of many, many tiny things (like rice), but as above it's a grammatical property: that's why even once you know a word is Count or Mass you can't be *sure* of the type of object it refers to. But you can still take a pretty good guess.

The properties of Count nouns are: they can co-occur with numerals, they can take “a”/“an” as an article, they co-occur with “fewer” but not “less” and “many” but not “much”, and you can't leave a singular count noun “bare” – that is, *without* an article (“the”, “a”/“an”), quantifier (like “some”, “every”), or numeral. Meanwhile, Mass nouns can occur “bare”, can't occur with numerals or “a”/“n” without a “measure” or “container” word like “grain”, “tablespoon”, “plate”, and co-occur with “less” but not “fewer” and “much” but not “many”.

In addition, some words act as “measures” or “containers” – they can take an “of <something>” phrase and, whether or not it's Mass or Count, turn it into Count. Words like these are necessary to use Mass nouns with numerals, “a”/“an”, etc.

How could you determine these properties in this problem if you didn't already know all this? Easy – put in words you do know in place of the unknown ones. For example, if a word like “water”, “rice”, “porridge”, etc. fits in the same places that “meembel” does and makes good English sentences, but not in the places “gelbelgarg” does, then it's very likely that “meembel” is something like water, rice, or porridge. Meanwhile, “burger(s)” fits in the same places “gelbelgarg” does, but not “meembel”, making it very likely that a “gelbelgarg” is some kind of discrete item.



REGISTRATION #:

(15 points)

(C) Lost in Yerevan (1/2)

On her visit to Armenia, Millie has gotten lost in Yerevan, the nation's capital. She is now at the metro station named **Shengavit** but her friends are waiting for her at the station named **Barekamutyun**. Can you help Millie meet up with her friends?



```

graph LR
    n((n)) --> a((a))
    a --> c((c))
    c --> l((l))
    l --> o(((o)))
  
```

(15 points)

(C) Lost in Yerevan (2/2)

C1. Assuming Millie takes a train in the right direction, which will be the first stop after Shengavit? Put the correct letter in the box on the right. Note that all names of stations listed below appear on the map. (4 points)

- a. Gortsaranayin
- b. Zoravar Andranik
- c. Charbakh
- d. Garegin Njdehi Hraparak
- e. none of the above

C2. After boarding at Shengavit, how many stops will it take Millie to get to Barekamutyun (don't include Shengavit itself in the number of stops)? (4 points)

C3. What is the name (transcribed into English) of the end station on the short, five-station line that is currently in construction, shown in a different shade on the map? Start writing from the leftmost box. (7 points)

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



2010 Solutions

(C) Lost in Yerevan (I/I)

This is a fairly simple problem to solve. One possible way to start is to look for the station on the map that has three words in its name. Then, based on its spelling as well as the spelling of the other given station names, one can easily determine that Armenian is written left to right and that both vowel and consonant sounds are represented as single characters in Armenian. Note that some sounds (e.g., "ts", map to multiple letters in English). After a few iterations, it is easy to reconstruct the entire part of the Armenian alphabet that is needed for this problem. The only "tricky" part has to do with the "T" character which doesn't appear in the names of the labeled stations, though it appears in the name of the subway system (METROPOLITEN).



(20 points)

(G) Tangkhul Tangle (I/2)

Tangkhul is a language spoken in the northernmost district of the Indian state of Manipur. Like Manipuri (or Meithei) and many other languages of Northeast India, Tangkhul is related to Tibetan and Burmese rather than to Hindi, Bengali, Marathi, Gujarati, or other well-known languages of India.

Tangkhul words can be very long and quite complicated in their structure. Sometimes single words may have to be translated with whole sentences in English. Also, pronouns (words like *he*, *she*, *it*, and *they*) can be left out if their meanings can still be filled in from context. Following are a list of sentences from Tangkhul and their English translations (in alphabetical order). In the English translations, pronouns are enclosed in parenthesis when they are left out of the Tangkhul sentences. Tangkhul, unlike Modern English (but like Old English), distinguishes three different grammatical numbers: singular (referring to one person or thing), dual (referring to two persons or things), and plural (referring to three or more persons or things). The abbreviations *sg.*, *dl.*, and *pl.* indicate “singular,” “dual” and “plural,” respectively.

GI. Match the Tangkhul sentences with their English translations by writing the number of the English translation by the corresponding Tangkhul sentence (8 points)

Tangkhul sentences

- a) a masikserra
- b) āni masikngarokei
- c) āthum masikngarokngāilā
- d) ini thāingarokei
- e) na thāilā
- f) ithum thāingāihāirara
- g) rāserhāira
- h) āni rāra
- i) nathum rāserhāiralā

English translations

- 1) Do they (pl.) want to pinch one another?
- 2) Do you (sg.) see it?
- 3) Have you (pl.) all come?
- 4) He/she will pinch all (of them).
- 5) (They) all have come.
- 6) They (dl.) pinched one another.
- 7) They (dl.) will come.
- 8) We (pl.) will have wanted to see (it).
- 9) We (dl.) saw one another.

A	B	C	D	E	F	G	H	I



(20 points)

(G) Tangkhul Tangle (2/2)

G2. Translate the following sentences into English. Always start with the leftmost box.
Please follow the style of the English translations given in G1 as closely as possible. (6 points)

a) nathum masikserngāira

b) āthum thāiei

c) i thāiserhāiralā

G3. Translate the following sentences into Tangkhul (6 points).

1) Do you (dl.) want to come?

2) You (sg.) have seen (it) all.

3) We (pl.) will want to see one another.



2010 Solutions

(G) Tangkhul Tangle (I/2)

Note that all but one of the Tangkhul sentences (sentence g) consist of two words. The two words consist of recurring components. For the first word, these are i, na, ā, ni, and thum. For the second word, these are masik, thāi, rā, ser, ngāi, ngarok, hāira, ei, lā, and ra. The word in the one word sentence (g) is drawn from components in the second set. In exactly one of the English sentences (sentence 5), the pronoun standing for the person doing the action of the verb is enclosed in parenthesis, showing that it is not present in the Tangkhul original. From this we may infer that:

1. sentences g and 5 match.
2. the final word in each Tangkhul sentence is the verb.
3. The first words in each two-word Tangkhul sentence must be a pronoun.

The pronouns vary in person and number. First person includes the speaker (I, we), second person includes the one being addressed (you), and third person refers to some other entity (he, she, it, they).

	sg	dl	pl	tot
1st	0	1	1	2
2nd	1	0	1	2
3rd	1	2	1	4
tot	2	3	3	8

Exactly one of the Tangkhul pronouns occurs twice (āni, in b and h). This must be 3rd dual. Therefore, these sentences must match sentences 6 and 7. The component ā occurs four times, like third person; the component ni occurs three times, like dual. Working in this direction, it is possible to establish the following equivalences:

1st i
2nd na
3rd ā

sg (unmarked)
dl ni
pl thum



2010 Solutions

(G) Tangkhul Tangle (2/2)

This establishes the following matches between the Tangkhul and English: a=4, c=1, d=9, e=2, f=8, i=3. Tangkhul sentences a and c both contain *masik*; the English equivalents both contain "pinch/pinched". d, e, and f all contain *thāi*; the English equivalents all contain "see/saw." Tangkhul sentences i and g both contain "*rā*"; the English equivalents both contain "come". Thus, the first part of the final word in the Tangkhul sentences is the verb root ("pinch", "see", or "come"). It follows that b=6 and h=7. Now that the sentences are matched, it is possible to determine the meanings of the verb suffixes:

ser all
ngarok reciprocal (X one another)
ngāi desiderative (want to X)
hāira perfective (have Xed)
ei past
ra future
lā interrogative

These are not problematic, except for *ser*. Transitive verbs are verbs which take object and intransitive verbs are verbs which do not. If *ser* is suffixed to the transitive verb *masik* "pinch", we get to "pinch all". That is, it quantifies over the object. However, if it is suffixed to the intransitive verb *rā* "see", we get "call come". That is, it quantifies over the subject. Given this observation, and the above equivalences, it is possible to provide the correct translations for G2 and G3.



Problem №5 (20 marks)

Given are pairs of cognate words of two closely related languages—Turkish and Tatar. Some words have been left out:

Turkish	Tatar	translation
bandır	mandır	dip!
yelken	cilkän	sail
onuncu	unıncı	tenth
baytar	baytar	vet
yiğirmi	yegerme	twenty
bencil	minçel	selfish
güreş	köräş	wrestling
işlesem	eşlăsäm	if I work
büyük	böyek	great
yıldırım	yıldırım	lightning
bunda	monda	in this, here
yetiştir	citeşter	convey!
göğër	kügär	become blue!
bozacı	buzaçı	<i>boza</i> handler
gerekli	kiräkle	necessary
boyun	muyın	neck
uzun	ozın	long
yöneliş	yünäleş	direction
	osta	master
	küzänäk	pore
	yılan	snake
yedişer		seven each
bilezik		bracelet
üstünde		on top of
bin		mount!
yumru		lump, swelling

§§. Fill the gaps.

Notes:

The letters ä, ı, ö, ü stand for specific vowels (the first two are not unlike the ones in *cat* and *bird*, respectively), while ğ is a specific Turkish consonant; c, ç, ş, y are pronounced as the initial consonants in *jet*, *chip*, *ship*, *yet*.

Boza is a weakly alcoholic drink made from millet.

Ivan Derzhanski

Editors:

Alexander Berdichevsky, Svetlana Burlak, Ivan Derzhanski, Dmitry Gerasimov (editor-in-chief),
Ivaylo Grozdev, Xenia Guiliarova, Boris Iomdin, Ilya Itkin, Axel Jagau, Alexander Piperski,
Maria Rubinstein, Michiel de Vaan

English text:

Alexander Berdichevsky, Ivan Derzhanski, Dmitry Gerasimov

Good luck!

Problem №5

By examining the data in the table we obtain the following correspondences between the sounds of Turkish and Tatar:

	Turkish		Tatar	note
1, 4, 11, 14	a	~	a	
2, 6, 12, 15	e	~	i	in the first syllable
7, 8, 13, 15, 18			ä	in a non-first syllable
3, 14, 16	o	~	u	
11, 17	u		o	in the first syllable
3, 16, 17			ı	<i>following u or o in Turkish</i>
1, 10, 14	ı			in a non-first syllable
13, 18	ö	~	ü	
7, 9	ü		ö	in the first syllable
9			e	<i>following ü {or ö} in Turkish</i>
5, 6, 8, 12, 15, 18	i			in a non-first syllable
5, 8, 10	m	~	m	word-medially
1, 6, 11, 16	b			word-initially if n follows somewhere
4, 9, 14			b	word-initially otherwise
1, 10, 11	d	~	d	
4, 12	t	~	t	
1, 2, 3, 6, 11, 16, 17, 18	n	~	n	
2, 6, 8, 10, 15, 18	l	~	l	
1, 4, 5, 7, 10, 12, 13, 15	r	~	r	
8	s	~	s	
14, 17	z	~	z	
7, 8, 12, 18	ş	~	ş	
3, 6, 14	c	~	ç	
2, 12	y	~	c	
4, 5, 9, 10, 16, 18			y	elsewhere
5, 13	ğ	~	g	
7, 13, 15	g	~	k	word-initially
2, 9, 15	k			elsewhere

Using these observations, we can reconstruct the missing words:

	Turkish	Tatar		Turkish	Tatar
19.	usta	osta	23.	bilezik	beläzek
20.	gözenek	küzänäk	24.	üstünde	östendä
21.	yılan	yılan	25.	bin	men
22.	yedişer	cideşär	26.	yumru	yomrı