ORMC Advanced 2: Linguistics III

Sunny Liang

Here are six linguistics puzzles arranged in approximate order of difficulty. Feel free to do the problems in any order as they interest you. Don't be afraid to collaborate with your peers and ask for help from instructors!

1 LOLWUT (NACLO 2017)

Recognizing words – like that the sequence of letters "c-a-t" represents the word "cat" – is the foundation to any technology that works on text... but it's not always so easy, especially on the internet where writers so often vary their spelling to express emphasis, emotion, surprise, etc. How can you recognize the word "what" when it could appear as "whaaaat" or "whaaa" or "wat" or "waaat" or "whut" or "wut" or...?

One way to recognize many variants at once is to use a *regular expression* (or *regex* for short) – a special sequence of symbols where you can indicate that a letter is optional, that it can occur many times, that one of several letters might occur, etc. For example, the regular expression

`wh?a+t*`

means that the "h" can occur exactly once or not at all, the "a" can occur one or more tiems, and the "t" can occur any number of times. The symbols used in this regex are defined as follows (in this example, *unit* refers to single letters, but as you will see below, a unit can be larger):

- `?` - the previous unit can occur zero or one times (but no more)
- `*` - the previous unit can occur zero or more times
- `+` - the previous unit can occur one or more times (but must occur at least once)

So the regular expression `wh?a+t*` would recognize

what wha whaaa wat waaaa watt waaaatttt waaaaat

but not

wht whut whhhat waaaah

There are many more symbols that can appear in regular expressions, but for this problem you only need the above three symbols and parentheses. Parentheses group letters into units that themselves can be operated on by symbols, so that `L(OL)+` would recognize

LOL LOLOL LOLOLOL LOLOLOLOL

but not

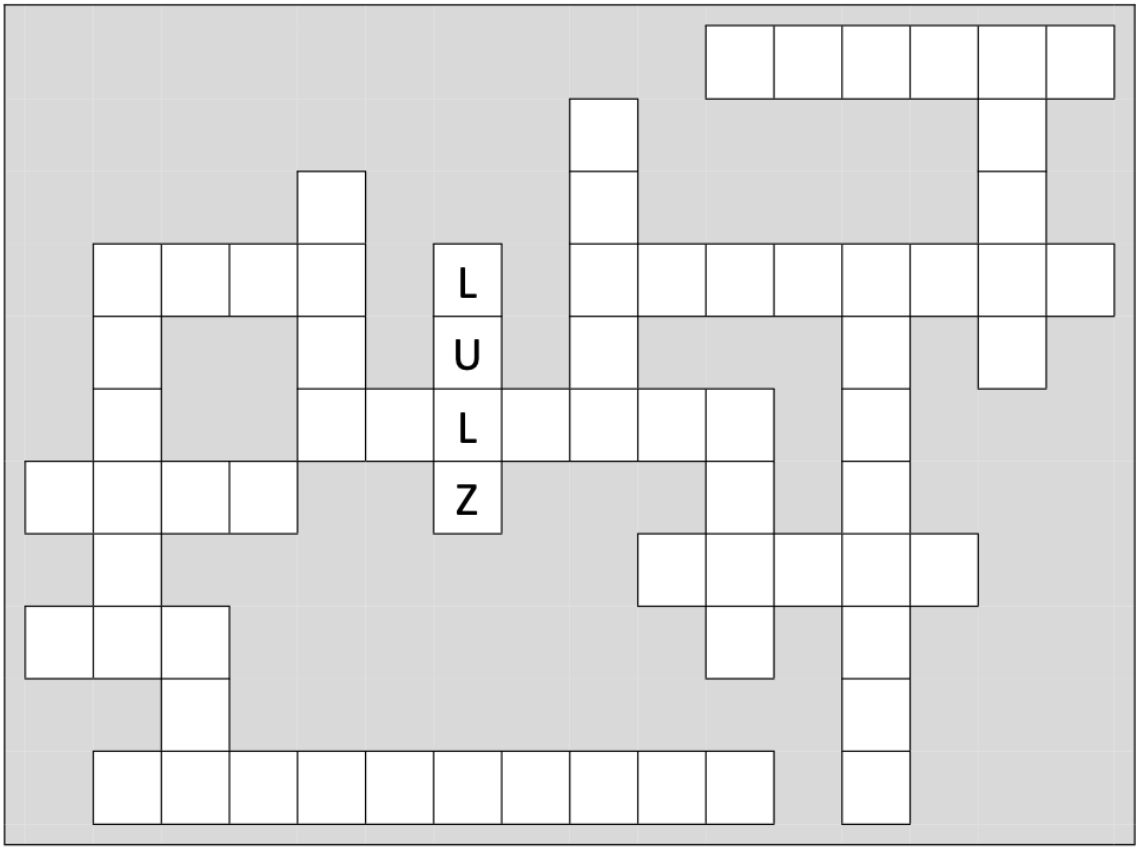
L LOOOOL LOOLL

For each of the following regexes, write the shortest word that it could describe:

- `LOL*LO?(OL)?O+L`
- `(A?HA)+(LO)?O*`

Here's a small crossword puzzle entirely of internet laughter, where each entry is clued with a regular expression like `L(OL)+`. The clues do not appear in any particular order, so you'll have to work out for yourself where each entry goes, but each clue corresponds to only one entry.

`L(OL)+`	`(HO)+`	`K(EK)*E`	`(HAR+)+`
`H(EH)+`	`ROT?FL`	`TE(HE+)+`	`LAW*L`
`MWA(HA)+`	`HE(HE)+`	`LO+L`	`HAHA*`
`(AH)+A+`	`HA+`	`(JA)+`	



2 Pidgin not Pigeon (All Ireland Linguistics Olympiad Sample)

When speakers of distinct languages which come into contact but share no common linguistic background need to communicate, they often create *pidgin* languages. Pidgin languages typically consist of linguistic features of the contact languages, and their vocabulary is typically derived from the socially or economically dominant language. A pidgin language for which English has supplied much of the vocabulary is Cameroonian Pidgin English (CPE) which is used in Cameroon, West Africa. The CPE examples here are represented phonetically. Note that *o* and *ó* (pronounced like the 'o' in *cot* and the 'aw' in *awful*, respectively) are different sounds.

For example, here are some CP words with their translations in English:

- *go* — 'go'
- *man* — 'man'
- *basiket* — 'basket'
- *haus* — 'house'
- *bet* — 'bed'
- *du* — 'do'
- *kol* — 'cold'
- *siton* — 'stone'
- *shet* — 'shirt'
- *do* — 'door'
- *dai* — 'die'/'dead'
- *fashon* — 'behaviour'
- *nes* — 'nurse'
- *sinek* — 'snake'
- *buk* — 'book'
- *sipun* — 'spoon'
- *kain* — 'kind'
- *simen* — 'cement'

It is typical for speakers in a complex multilingual situation, as is the one which creates pidgin languages, to find creative ways to extend the usefulness of a small borrowed vocabulary. One way to accomplish this goal is by metaphorically extending borrowed words, as in the examples below, which are CPE compounds derived metaphorically from the English word *eye*.

- *ai kón* — 'ear of corn'
- *benben ai* — 'cross-eyed'
- *lóng ai* — 'covetous'
- *bat ai* — 'evil eye'
- *bik ai* — 'greedy'
- *tróng ai* — 'courageous'

On the next page, you are presented with CPE vocabulary items derived in a similar fashion from the English words as shown. Your task is to match the CPE compounds given for each word to their likely English translations on the right.

Word	***Phrase***	**Meaning**
han 'hand'	*open han*	one of a pair
	wan han	generous
	tai han	mean
het 'head'	*bik het*	stubborn
	klin het	innocent
	tróng het	intelligent
hat 'hat'	*bat hat*	kind/generous
	blak hat	calm
	gut hat	wicked
	kul hat	clever/crafty
	kóni hat	hated/hateful
maut 'mouth'	*bik maut*	ability to persuade/flattery
	kóni maut	haughtiness/bragging
	swit maut	dishonesty
blak 'black'	*blak ai*	wicked
	blak bush	heart of a forest
	blak hat	dizzy
drai 'dry'	*drai ai*	unsubstantiated statements
	drai han	thirst
	drai nek	poverty
	drai tók	insomnia

Furthermore, answer the following:

- There are two different ways to say 'generous' in CPE. What are they?
- What do you think the following CPE words are in English?

tróng *klin* *tók* *swit*

- Groups of consonants (called 'consonant clusters') in words borrowed from English are preserved in some contexts but simplified in others. Based on the words you have seen so far, predict what the CPE versions of the following English words will be.

stick *old* *green* *ground*

3 Babylonian Cuneiform Numerals (Princeton Linguistics Challenge 2005)

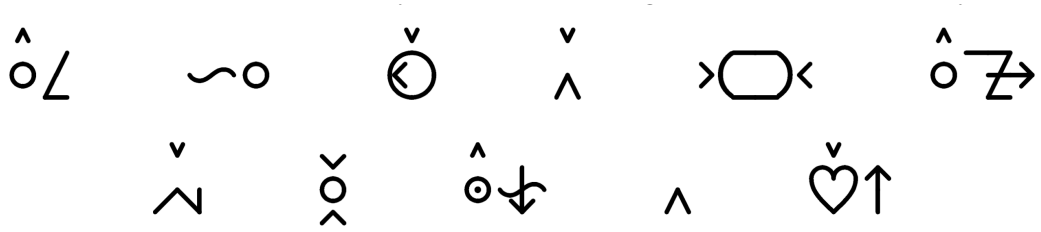
The earliest known writing system originated over 5000 years ago in what is now Iran, Iraq and other parts of Western Asia. This writing system, called *cuneiform*, was used by the ancient Persian kings to make their decrees known, and to audit the tax returns of their many subjects. The characters were inscribed on clay or stone tablets using wedge-like instruments. Although many inscriptions have survived, the writing system was not deciphered by modern scholars until 1846.

In this puzzle you will carry out the kind of work that these scholars had to do to decipher the cuneiform writing system. The following is a fragment from a Babylonian educational document that was discovered in 1811. It was this document that allowed scholars to unlock the number system used by the ancient Babylonians. From this, scholars were able to extend their understanding to the entire writing system. Many of the characters are illegible because of the ravages of time. Nevertheless, it is possible to figure out what the missing characters should be. Your job is to fill in the missing characters:

4 Blissymbolics (IOL 2010)

Blissymbolics is a universal system of symbols devised by Charles K. Bliss (1897–1985), an Australian of Austrian origin, who thought it should be understandable to all people, regardless of their native tongue. To this day, people with severe speech and physical impairments often use Blissymbols to help them communicate.

Given are words written in Blissymbolics and their English translations in arbitrary order:



waist; active; ill, sick; lips; activity; to blow; western; merry; to weep; saliva; to breathe.

- Determine the correct correspondences.
- Indicate what the following symbols mean. *Hint*: Two of them have the same meaning.

- Write in Blissymbolics: air, body (torso), to rise, east, sad.

5 Icelandic Kinship (Brazilian Linguistics Olympiad 2013)

In Iceland, last names don't work the way they do here. See if you can figure it out!

Guðrún Eriksdóttir Hrafnhildardóttir and Jakob Christiansson had three children who have all had children of their own. In fact, some of Guoron and Jakob's grandchildren have had their own children as well. In order to celebrate their 70th wedding anniversary, the long-lived couple had a party with all their descendants: sons and daughters, grandsons and granddaughters, great-grandsons and great-granddaughters. Naturally, some of their sons-in-law and daughters-in-law, as well as grandsons-in-law and granddaughters-in-law, were also invited. Here is the list of all the people in the party:

- Christian Kristínsson
- Daniel Guðrúnarson Jakobsson
- Daniel Steinunnarson Đorarinssonar
- Eva Emilsdóttir Kvaran
- Gunnar Gunnarsson
- Hrafn Kristínsson Viktorsson
- Ingimundur Sigurðarson Bergmann
- Jakob Đorarinsson
- Jón Oddsson Bergmann
- Kristín Guðrúnardóttir Aronsdóttir
- Margret Steinunnardóttir Jakobsdóttir
- Ragnheiður Jakobsdóttir
- Rakel Ragnheiðardóttir Bergmann
- Robert Bergmann Gunnarsson
- Sara Jakobs Đorarinssonar
- Sigurður Jóns Bergmann
- Stefan Gunnarsson Gunnarssonar
- Steinunn Jakobsdóttir
- Viktor Danielsson Jakobssonar
- Viktoría Viktorsdóttir

Note: **j** is pronounced as *y* in *you*; **Đ** and **ð** are pronounced as *th* in *this*.

1. Some members of the party are descendants of foreigners. Which is the oldest?
2. Which of the children of Guðrún and Jakob has no grandchildren?
3. How many children does Viktor have?
4. How many children does Steinunn have?
5. Write five possibilities of name and surname for the son that Eva and Christian are awaiting, knowing that his first name will be the same as the father of the pop singer Björk Guðmundsdóttir.

6 Teeline (ORMC 2025)

Before the advent of audio recording, the only way of recording things that were said out loud was by writing them down. Courts, for example, would have professional *stenographers* whose job it was to write down everything that was said during a trial. Writing is typically a lot slower than talking, hence, *shorthand* was invented to help stenographers keep up.

There are many popular shorthand systems, this problem deals in particular with *Teeline*. While most shorthand systems are phonetics-based (that is, based on pronunciation), Teeline is (mostly) spelling-based. All shorthand systems are fairly flexible and different stenographers will come up with their own techniques and modifications to a given system.

Below are two pangrams written in Teeline by a computer program. Actual stenographers would probably write these differently, but these are good enough for our purposes.

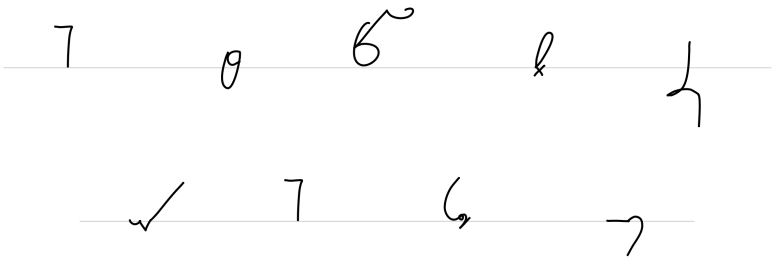
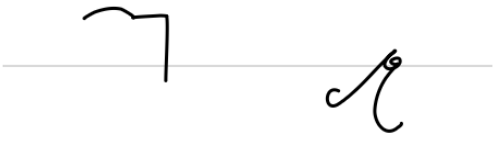


Figure 1: the quick brown fox jumps over the lazy dog

Figure 2: pack my box with five dozen liquor jugs

What does the following say?



Write the following sentence in Teeline:

Everybody is identical in their secret unspoken belief that way deep down they are different from everyone else.