ORMC Advanced 1: Linguistics

Sunny Liang & Curtis Liu

May 2024

This handout consists of four problems adapted from the International Linguistics Olympiad followed by five problems that are more math-oriented. Feel free to do the problems in any order as they interest you. Some problems are not easy to do on your own, you are encouraged to collaborate with your peers and ask for help from instructors!

1 Georgian Countries

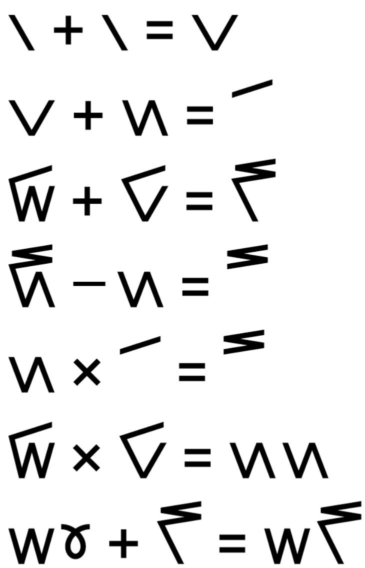
Here are the names of five countries written in Georgian, a 33-letter alphabet. The English translations are provided for three of them. What are the other two?

ბრაზილია	Brazil
პერუ	Peru
ურუგვაი	Uruguay

არგენტინა
კოლუმბია

2 Kaktovik numerals

Many of the languages spoken by the indigenous peoples of Alaska and Northern Canada use a base-20 number system. In 1994, students in Inuktitut, Alaska invented a base-20 numeral system that more naturally corresponded to their Inupiaq language than our base-10 Arabic numerals. Below is a series of arithmetic operations using Inuktit numerals. Can you determine all 20 symbols representing the numbers from 0 to 19?[1]



Using *positional notation*, the Inupiaq people can represent all the integers larger than 19 without extra symbols. We also use positional notation with Arabic numerals to represent all the integers larger than 9. Given your understanding of how positional notation works with Arabic numerals, figure out the Inuktitut representations for the following numbers:

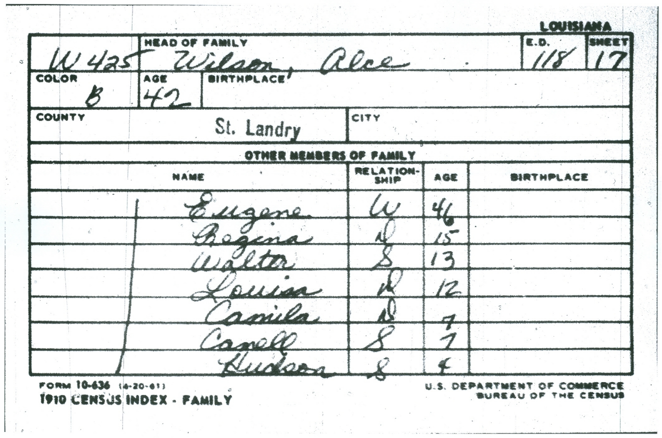
- 20
- 43
- 1234
- 0.000375

[1] Hint: Owing to a long and messy history, Arabic numerals have rather arbitrary shapes. Inuktitut numerals, on the other hand, are highly systematic and follow very logical patterns. Noticing these patterns will help you solve this problem.

3 Soundex

Soundex is an algorithm for indexing names. Many surnames sound similar but vary slightly in spelling, Soundex tries to group names together that sound similar but may not look similar. It was invented in the early 20th century to make it easier to search for surnames. Most notably, it was used extensively to help analyze US census data. Today, modernized variants of Soundex are used in database software such as SQL.

Below is a sample card from the 1910 census. It belongs to Alice Wilson and is indexed with Soundex code W425.



LOUISIANA

	HEAD OF FAMILY	E.D.	SHEET
W425	Wilson, Alce	118	17

COLOR	AGE	BIRTHPLACE
B	42	

COUNTY	CITY
St. Landry	

OTHER MEMBERS OF FAMILY

NAME	RELATIONSHIP	AGE	BIRTHPLACE
Eugene	W	46	
Regina	D	15	
Walter	S	13	
Louisa	D	12	
Camila	D	7	
Carell	S	7	
Hudson	S	4	

FORM 10-636 (4-20-61)
1910 CENSUS INDEX - FAMILY

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

Here are some more examples of Soundex encodings:

- Allaway: A400
- Buckingham: B252
- Chapman: C155
- Carraway: C600
- Daniel: D540
- Daniels: D542
- Descartes: D263
- Euclid: E243
- Hilbert: H416
- Robert, Rupert: R163
- Tao: T000
- Kafka: K120

Try to determine the steps of the Soundex encoding algorithm. Which letters are ignored? Which letters produce the same digits? What about repeated digits? Once you have the algorithm, generate Soundex codes for the following surnames:

Noether, Hypatia, Ramanujan, Bernoulli, Schrodinger, your own surname.

4 Toki Pona

Toki Pona is a sort of experimental language devised by Canadian Linguist Sonja Lang in 2001. It subverts the abstractions of modern languages by being utterly minimal and concrete. It consists of just 14 phonemes (sounds) and about 120 essential words.

Here are 20 phrases in Toki Pona:
kiwen suno jelo, tomo tawa telo, jan poka, ilo suno, telo jelo, jan ilo, jan toki, supa lape, supa moku, ma tomo, wile moku, tawa, nasin linja, wile pona, telo kiwen, lipu toki, wile lawa, linja lawa, tomo moku, linja kiwen

Here are their unordered English translations:
prophet, well-intentioned, hair, lantern, ice, robot, boat, thorn, hungry, friend, book, pee, city, bed, orthodoxy, movement, restaurant, dominant, dinner table, gold

- Pair up the English translations to their Toki Pona counterparts. Try to do this systematically by mapping out how the words might relate to each other.
- What do each of the 'essential' words here mean? For example, you could say that *jan* means *person*.
- What does the name of the language, Toki Pona, mean?

For the next problem, you need to know some basic group theory. A *group* in mathematics is a non-empty set with a binary operation that takes two elements of the set to produce another element of that set. A group must also satisfy the three *group axioms*. Let G be a group. We will represent its elements with lower-case letters and its operation with a dot $\cdot$. A group G must satisfy:

1. Associativity: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
2. Identity: there is some $i \in G$ such that $i \cdot a = a$ for any $a \in G$
3. Inverse: For every a, there exists some unique b such that $a \cdot b = b \cdot a = i$ where i is the identity.

For example, the set of integers with the operation $+$ form a group. Integer addition is associative. The identity is 0. Lastly, every number has a unique inverse, namely its negative.

5 The Homophonic Group

By definition, English words have the same pronunciation if their phonetic spellings in the dictionary are the same. The homophonic group $\mathcal{H}$ is generated by the letters of the alphabet, subject to the following relations: English words with the same pronunication represent equal elements of the group. Thus, *be* = *bee*, and since $\mathcal{H}$ is a group, we can cancel *be* to conclude that $e = 1$. Try to determine the group $\mathcal{H}$. Don't let the jargon intimidate you, this problem is easier than it seems!

6 Syllables and Factors

When written as an English word, how many of the first 100 positive integers have as many factors as syllables?

7 Deciphering Japanese

Below is part of a multiplication table in random order and in Romanized Japanese. Fill in the question marks with the appropriate values.

futatsu × yottsu = yattsu	itsutsu × itsutsu = ni juu go
yattsu × kokonotsu = nana juu ni	itsutsu × yattsu = yon juu
mittsu × mittsu = kokonotsu	kokonotsu × mittsu = ni juu nana
muttsu × mittsu = juu hachi	futatsu × rei = rei
kokonotsu × ? = haci juu ichi	muttsu × kokonotsu = ?
yottsu × ? = san juu ni	futatsu × nanatsu = ?

8 Recurrence Language

A word is a string of the symbols a, b which can be formed by repeated application of the following:

1. ab is a word;
2. if X and Y are words, then so is XY;
3. if X is a word, then so is aXb.

How many words have 12 letters?

9 Microbablia

In the land of Microbablia, the alphabet has only two letters, 'A' and 'B'. Not surprisingly, the inhabitants are obsessed with the band ABBA. Words in the local dialect with a high ABBA-factor are considered particularly lucky. To compute the ABBA-factor of a word you just count the number of occurrences of ABBA within the word (not necessarily consecutively). For instance, AABA has ABBA-factor 0, ABBA has ABBA-factor 1, AABBBA has ABBA-factor 6, and ABBABBA has ABBA factor 8. What is the greatest possible ABBA-factor for a 100 letter word?