P#1.1: Tonight or Tomorrow?

In the following Table, there are 16 different spellings that people typed for the words “tomorrow” and “tonight” in English Tweets; 8 of these correspond to “tomorrow” and another 8 to “tonight”. However, due to some encoding error, the Latin characters are all mapped to some other symbols. Your task is to recover the original words (write them down in the space provided under Latin Transcription).

Code Word	Latin Transcription	Code Word	Latin Transcription
🗎❍❑❒❑		♦❑■🖂♦	
♦❑❍❑⌘		🗎■♓♑♒♦	
♦❑❍❑❒❑		🗎■🖂♦	
♦❑❍❒⬥		🗎■🖰	
♦❑❍		♦❑■🖰	
♦❑❍❒♋		♦❑■🖂♦⌘	
🗎❍❒❑		🗎■♓♦♏	
❍❑❒♋		♦❑■♓♦	

Now decode the following Tweet (into Standard English):

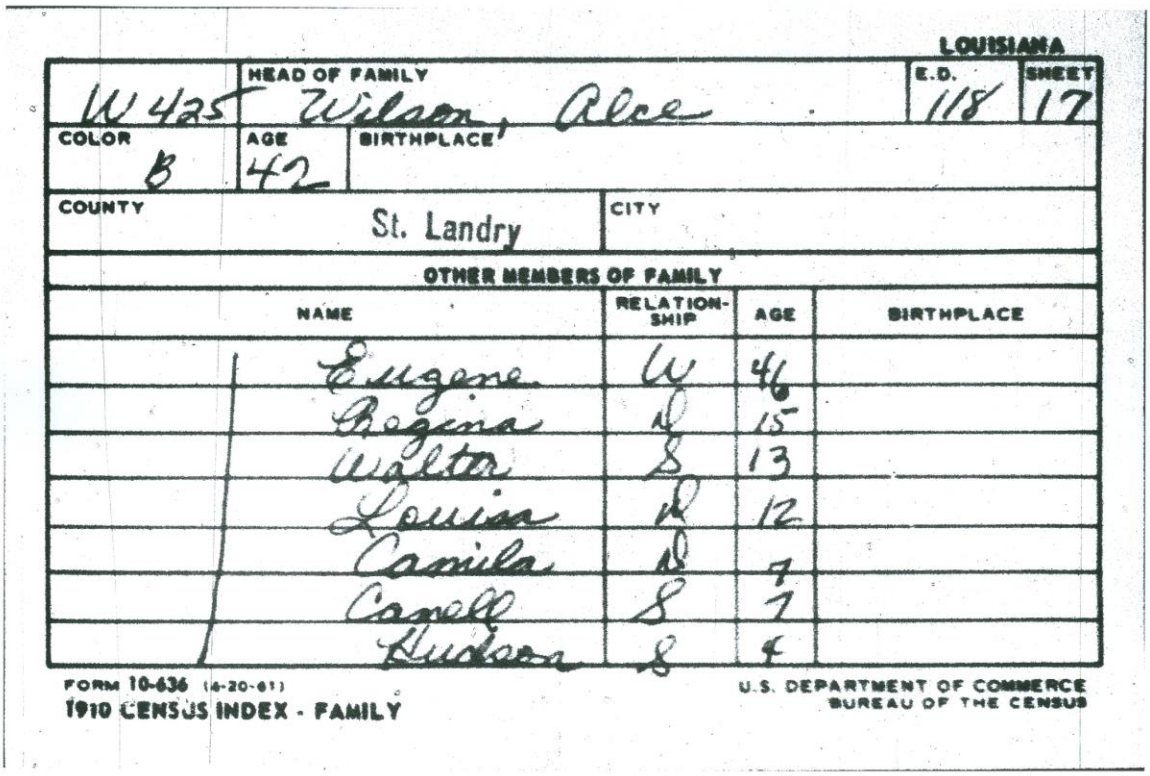
❒🖂♦ ■⬥ ⬥♏ ❒ ♎♏♍❑♎♓■♑ ⬥♓■♑♎♓■♑⌘

Answer:

_____

P#1.2 SOUNDEX

Soundex is an algorithm for coding names. It was developed in the USA in 1918–1922 by Robert C. Russell and Margaret King Odell in order to facilitate searching for similar-sounding surnames. In the middle of the 20th century, Soundex was extensively used in the USA to analyze results of 1890–1920 censuses. Below is a sample card with data from the 1910 census. You can see the Soundex code for *Wilson*, which is W425.



LOUISIANA

	HEAD OF FAMILY	E.D.	SHEET
W 425	Wilson, Alce	118	17

COLOR	AGE	BIRTHPLACE
B	42	

COUNTY	CITY
St. Landry	

OTHER MEMBERS OF FAMILY

NAME	RELATION-SHIP	AGE	BIRTHPLACE
Eugene	W	46	
Regina	D	15	
Walter	S	13	
Louisa	D	12	
Camila	D	7	
Carroll	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 is a list of surnames, with the corresponding Soundex codes in arbitrary order. Some characters are missing:
Allaway, *Anderson*, *Ashcombe*, *Buckingham*, *Chapman*, *Colquhoun*, *Evans*, *Fairwright*, *Kingscott*, *Lewis*, *Littlejohns*, *Stanmore*, *Stubbs*, *Tocher*, *Tonks*, *Whytehead*

S312, T␣6␣, ␣5␣3, C42␣, T520,

L␣42, A536, C155, ␣623, S356,

␣252, ␣152, ␣330, A251, A400, L2␣0

(a) Match the surnames with the corresponding Soundex codes and restore the omitted characters.

(b) Give the pseudocode/algorithm for converting an English word to its Soundex code.

(c) Generate Soundex codes for the following surnames:
Ferguson, *Fitzgerald*, *Hamnett*, *Keefe*, *Maxwell*, *Razey*, *Shaw*, *Upfield*.

—problem designed by *Alexander Piperski*
(Appeared in Int. Linguistics Olympiad, 2015)

P#2.2: Māori Loanwords

The Māori language, or "te reo Māori", is the language of the Māori, the indigenous people of New Zealand. It is one of the official languages of New Zealand, along with English and New Zealand Sign Language, and over several centuries it has borrowed many English words. These words are often adapted to better fit the sounds of the Māori language.

2.2A Below are 20 Māori words that have been adapted from English words. Note that Māori uses a line over vowels to mark them as long. Can you match each word below to the picture that illustrates it?

hāma	*māti*	*raina*	*tīhi*
hāpa	*paipa*	*taraka*	*tūru*
hū	*piriti*	*terewhono*	*wāna*
hūtu	*pūnu*	*tiā*	*whurutu*
iniki	*pūtu*	*tiaka*	*wūru*

2.2B. Many English loanwords in Māori deal with introduced Western professions and titles. To what English words do the following Māori words correspond?

hekeretari, pirinihehe, pirihimana, tiati

2.2C. What countries are these? *Iharaira, Kiupa, Peina, Tiamani, Tiapana*

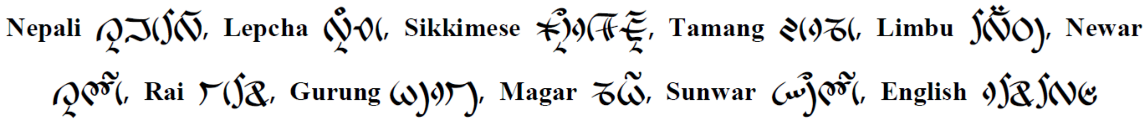
2.2D. For each of these English words, predict what the Māori form would be:

beef, bull, cart, clock, lease, meat, seal, street, time, watch

(Appeared in North American Computational Linguistics Olympiad 2014)

P#2.2 Transliterating Lepcha

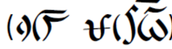
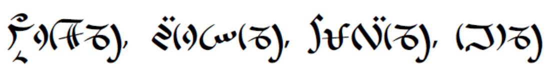
Sikkim has 11 official languages, which are listed below in English as well as the *Lepcha Script*.



Assignment 1:

The *Lepcha* name for one of the above languages does not match its English name. Which one? It is actually written as **Drendzongke** (*dz* is pronounced somewhat like the *j* in jug.)

Assignment 2:

The *Lepcha* speakers, who call themselves the **Roong haagiit** (), are composed of four main distinct communities: 

Transcribe these four community names into English.

Assignment 3:

Sikkim boasts of the third highest mountain peak of the world, *the Kangchenjunga*. Transcribe this name in the Lepcha script, considering the fact that it is pronounced as **Kaang-chen-dzong-gaa** (means "*the five treasures of the high snow*" in the Tibetan language).

Note: The *Lepcha* script, used for writing the *Lepcha* language, is derived from the *Tibetan* script, and may have some Burmese influence. Around 30 to 50 thousand people speak the *Lepcha* language, who are spread across Sikkim, Northern West Bengal, Bhutan and Nepal.

- Problem designed by Monojit Choudhury

(Appeared in Panini Linguistics Olympiad for Juniors, Round 1, 2015)

P#2.3 Switching *ya* Mixing?

Code-Switching and *Code-Mixing* are typical phenomena of multilingual societies. Linguists differentiate between the two, where Code-Switching is juxtaposition, within the same speech exchange, of passages of speech belonging to two different grammatical systems (i.e., languages), and Code-Mixing refers to the embedding of linguistic units such as phrases and words of one language into an utterance (i.e., clause or sentence) of another language. This two different phenomena are illustrated using two comments taken from Facebook. Example A features Code-mixing where English words are embedded in a Hindi sentence, whereas Example B shows Code-Switching between English and Hindi.

Each examples are further split into sentences or clauses and marked using the tag <matrix= Language>… </matrix>. The *matrix* language is defined as the language which governs the grammatical relation between the words or phrases of the utterance. Any other language words that are nested into the matrix constitute the embedded language(s).

Note: Knowledge of Hindi is not necessary to solve this problem. Nevertheless, the meanings of the Hindi words are provided in square brackets and English translation of the entire sentence is also provided.

Example A: Code-mixing

> <matrix=Hindi> Love affection lekar [having carried] salose [for years] sunday ke [of] din [day] chali [run] aa [come] rahi [continuous tense marker] divine parampara [tradition] ko [object case marker] aage [forward] badhaa [stretch] rahe [continuous marker] ho [is]. </matrix>

Translation: With love and affection, [you] have been carrying forward the divine tradition that has been running on Sundays for years.

Example B: Code-switching

> <matrix=English>Great news! </matrix> <matrix=Hindi>is [this] bat [note] par [on] to [grammatical particle] ek [a] party [party] honi [be] chahiye [should]. </matrix> <matrix=English>Let's know </matrix><matrix=Hindi> ki [that] kab [when] aaye [come] tumhare [your] ghar [house]. </matrix>

Translation: Great news! There should be a party on this note. Let's know when shall [I/we] come to your house.

For each of the following Facebook posts, indicate whether they are cases of code-mixing or code-switching or both. Split the posts into matrices and indicate the matrix language using the same formatting style as shown in the examples above.

2.1a Dude I think u should try again caz ye [this] tera [your] fault nahi [not] hai [is]. ye [this] CBSE walo [people] ki [of] fault hai [is].

Translation: Dude I think you should try again because it is not your fault at all. This is CBSE folk's fault.

Answer:

2.1b Corruption to [grammatical particle] every level pe [at] hai [is] and its complete eradication possible nahin [is not].

Translation: Corruption is at every level and its complete eradication is not possible.

Answer:

2.1c I had told you exams me [in] difficult questions aayenge [will come].

Translation: I had told you that there will be difficult questions in the exam.

Answer: