

A Linguistics Challenge

Aalok Thakkar

December 7, 2015

Abstract

It is not a matter of surprise that *Pāṇini* in his *Aṣṭādhyāyī* presents to us a first of its kind insight into formal mathematical logic, along with the marvels of Sanskrit grammar and linguistics. Since then, mathematics and linguistics, and mathematicians and linguistics have had a close relationship. Certainly, there are no second order differential equations that help classify consonants based on their phonetics, nor do we find the solution to Diophantine problems using cognitive linguistics, however, there is a fundamental sense of logic and pattern matching that unite the two disciplines. The following challenge endeavors to help the readers appreciate this fact.

1 Agta

The following list of words is from the Agta language of the Central Cagayan Valley in the Northern island of Luzon, in the Philippines. There are now only about 600 speakers of this variety of Agta, although there are perhaps 10,000 people in the Philippines who speak other varieties also known as Agta. The Agta people now speak an Austronesian language similar to other languages spoken in the Philippines. However, they are descended from the Melanesian people who were present in the Philippines before the Austronesian peoples arrived. The Agta language is now seriously endangered. This question is adapted from Healey, Phyllis M. 1960, *An Agta grammar* (Manila Bureau of printing).

| English | Agta |
|------------------|-----------|
| creek | wer |
| little pig | balabahuy |
| beetle | talobag |
| granny | bakbakat |
| little money | palapirak |
| pig | bahuy |
| loincloth | bag |
| little creek | walawer |
| little loincloth | balabag |
| leg | takki |
| patch | labang |

Given this information, translate the words 'little leg,' 'money,' 'little beetle,' 'little patch' and 'little granny' to Agta.

2 Czech

The second problem is in the same lines as the first one and deals with pattern matching. Czech (formerly known as Bohemian), is a West Slavic language spoken by over 10 million people. It is the official language in the Czech Republic, and has minority language status in Slovakia. Czech's closest relative is Slovak, with which it is mutually intelligible. It is closely related to other West Slavic languages, such as Silesian and Polish, and more distantly to East Slavic languages such as Russian. The following are examples of how to say in Czech what time it is:

| Time | English translation | Czech translation |
|------|--------------------------|-----------------------------|
| 7:55 | Five minutes to eight | Za pet minut osm |
| 7:50 | Ten minutes to eight | Za deset minut osm |
| 7:15 | Quarter past seven | Ctvrť na osm |
| 7:52 | Seven minutes to eight | Za sedm minut osm |
| 7:07 | Seven minutes past seven | Za osm minut čtvrt na osm |
| 6:05 | Five minutes past six | Za deset minut čtvrt na osm |
| 7:30 | Half past seven | Půl osmé |
| 8:30 | Half past eight | Půl deváté |
| 5:20 | Twenty past five | Za deset minut půl šesté |
| 9:15 | Fifteen past nine | Ctvrť na deset |

This information should help you translate ‘twenty three minutes past five’ and ‘ten minutes past nine’ in Czech. Try your hand at it.

An interesting thing about Czech (and other languages with Slavic roots) is that the pronunciations are drastically different compared to other Indo-European languages. Check it out!

3 Binary Tree

In the formal study of sentence syntax, sentences are often represented as hierarchical trees. Here is a tree for the sentence “My dog ate an old hamburger.”

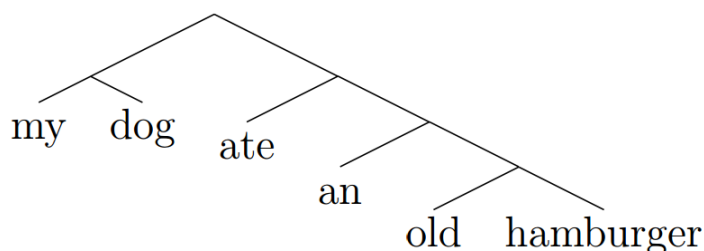


Figure 1: My dog ate an old hamburger

We shall call this sentence (1). Each vertex of a tree has either one word or two vertices below it. The tree above has 11 vertices.

Define a subsentence to be a continuous subset of words from a sentence.

Subsentences of (1) include, for example, “my dog ate an old hamburger,” “my dog ate,” and “old”. Strings which are not subsentences of (1) include “ate hamburger,” “old ham,” and “hot dog.”

Define a constituent to be a subsentence for which there is a vertex in the tree which contains (below it) that subsentence and no other words.

Constituents of (1) include, for example, “my dog,” “an old hamburger,” and “an.” Subsentences of (1) which are not constituents include “my dog ate” and “an old.”

Suppose a non-null subsentence is chosen randomly from an N-word sentence. Compute in terms of N the probability that this subsentence is a constituent.

4 Guarani

This question deals with an aspect of linguistics, known as typology. Guarani is an indigenous language of South America and is one of the two official languages of Paraguay. This, unlike the former two is an agglutinative language, that is words are formed by combination of smaller morphemes (grammatical

The following Guarani verb forms are listed along with their English translations.

| | | | |
|------------------------|------------------------|----------------------|-----------------------|
| <i>japyhyta</i> | We will be catching | <i>noñe'ëi</i> | He is not talking |
| <i>nohyvykõiri</i> | He is not enjoying | <i>okororõ</i> | He is crying |
| <i>ombokapu</i> | He is shooting | <i>ndajajupirima</i> | We were not waking up |
| <i>pemomaitei</i> | You are greeting | <i>ahyvykõima</i> | I was enjoying |
| <i>ndokarumo'ãi</i> | He will not be eating | <i>añe'ëta</i> | I will be talking |
| <i>ndapevo'oima</i> | You were not taking | <i>namomaiteiri</i> | I am not greeting |
| <i>napekororõmo'ãi</i> | You will not be crying | <i>japurahei</i> | We are singing |

1. Translate into English.

- a) *akaruma* _____
- b) *ojupita* _____
- c) *ndavo'omo'ãi* _____
- d) *napekororõi* _____
- e) *ndapyhyima* _____

5 Basque Numbers

Archaeologists, Cryptographers, Historians and Linguistics often deal with real world situations where it is impossible to translate a group of words from the given language. The only information available is of how a particular word or set of words behaves with another word or set of words. Such is the last question. Given below are mathematical (simple arithmetic) operations on whole numbers in Basque language. Try to deduce which numbers do each word stand for.

(Please put multiplication symbol instead of '*)

$$\text{azpiko} \times \text{bi} = \text{azpiko}$$

$$\text{bi} \times \text{bi} = \text{lau}$$

$$\text{bi} + \text{lau} = \text{sei}$$

$$\text{bat} + \text{bi} = \text{hiru}$$

$$\text{hiru} + \text{bat} + \text{bi} = \text{sei}$$

$$\text{bat} \times \text{bat} = \text{bat}$$

(Note: Each word represents a unique number, and each number is represented by a unique word)