Gil Caplan – HW1 Language Computation

Q1 Solution:

I will divide my life into sections as I am a native English speaker and moved to Israel at the age of 8.

Until I moved to Israel at the age of 8, I was exposed to a negligible number of Hebrew tokens and as such I will not include it in the count. The only word I knew in Hebrew when I moved to Israel was cucumber ("מלפפון").

After my family and I moved to Israel, I started ulpan and 4rth grade, let's split the week to the school week and the weekend. At that stage of life, let's assume I was exposed to 3000 tokens per hour in Hebrew when at school or activities not at home as we spoke English at home. 7 hours of Hebrew on schooldays and 5 hours on the weekend

Therefore, 4 years of primary school comes out to 208 weeks => 24*208*3000

Approx. 15 million tokens

Then 6 years of middle & high school. Let's assume I was exposed to 5000 tokens per hour at this stage with 7 hours of school on a school week and 7 hours over the weekend (including social activities)

Therefore, 6 years is 312 weeks => 312*24*5000.

Approx. 37.5 million tokens

Followed by 2.5 years of university, Let's assume I was exposed to 6000 tokens per hour. The timing is a bit more flexible in terms of what I'm doing so let's assume 60 hours a week on average. 2.5*52= 130 weeks => 312*24*6000

Approx. 45 million tokens

In addition, as I'm a religious Jew from the age of 13 I have been praying every morning and therefore, exposed to tokens in Hebrew. Let's assume 4200 tokens per hour where on avg 1.5 hours each day.

52 * 11 * 7 = 4,004 hours => 4,004 * 4200 tokens

Approx. 17 million

In aggregation, I was exposed to approximately 114.5 million tokens

Q2 Solution:

P(possessor is postnominal) = 1 – P(possessor is prenominal)

= 1 – (P(possessor is prenominal| possessor is animate)*P(possessor is animate) + P(possessor is postnominal| possessor is inanimate)*P(inanimate))

= 1 – (0.9 * 0.4 + 0.25 * 0.6) = 0.49

Bayes: P(possessor is animate| possessor is postnominal)

= (p(possessor is postnominal| possessor is animate) * P(possessor is animate)) / P(possessor is postnominal) =

[(1 - p(possessor is prenominal| possessor is animate)) * P(possessor is animate)]/ P(possessor is postnominal)

= (0.1 * 0.4) / 0.49 = 0.0816

Q3 Solution:

Entropy:

$$\mathrm{H}(W) = -\sum_{i=1}^{4} P(x_i) \cdot \log_2\big(P(W_i)\big) = \frac{1}{2}\log_2\frac{1}{2} + \frac{1}{4}\log_2\frac{1}{4} + \frac{1}{8}\log_2\frac{1}{8} + \frac{1}{8}\log_2\frac{1}{8}$$

$$= -\frac{1}{2} - \frac{1}{2} - \frac{3}{8} - \frac{3}{8} = -1.75 \textit{ (bits)}$$

Maximum is when all the different possibilities have the same probability to occur i.e. uniform across all options.

$$\mathrm{H}(W) = -\sum_{i=1}^{4} P(x_i) \cdot log_2\big(P(W_i)\big) = -\frac{1}{4}\log_2\frac{1}{4} \cdot 4 = 2$$

Q4 Solution:

The experiment is problematic since the setup of the experiment doesn't consider the effects of noise/interference (electromagnetic or background noise) which can have a profound effect on the study.

We can also notice that the integrity of the semantic incongruity between the two words that are changed to "drink" & "transmitter" makes them independent of each other which makes it hard to compare. This can be noticed as the words sound very different.

A solution that can be implemented is to address both these problems. Rather than using the word "transmitter" perhaps use a word that sounds like drink but has a different meaning. As well as using a Faraday cage to minimize noise and electromagnetic disturbances.