

Psych156A/ Ling150
Spring 2012
Homework 2: Words
Or “The Language Adventures of Sigmund von Hacklestein, part 2”

Remember to write your full name and University ID number on your assignment. If you collaborate with other students in the class, please make sure to indicate who you worked with.

(65 points total)

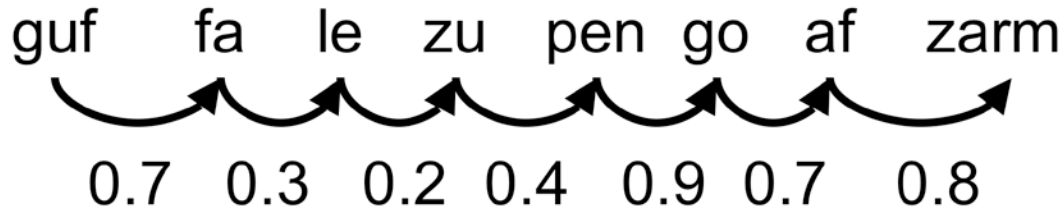
(1) Sigmund has brought back some information about the language of Guin. Here are all the words he found that begin with “pen”, separated by syllables (there is a / between syllables):

pen/bat	=	“penbat”
pen/ga/mor	=	“pengamor”
pen/ta/pen/bat	=	“pentapenbat”
pen/ma/kist	=	“penmakist”
pen/ga/kist/ly	=	“pengakistly”
pen/go/ta/kist/ly	=	“pengotakistly”
pen/kist	=	“penkist”
pen/bat/kist/ly	=	“penbatkistly”
pen/ga/mor/ly	=	“pengamorly”
pen/lo/mor/ly	=	“penlomorly”

Calculate the transitional probability of the syllable sequences below, given these data. You should write your answer as # / # (ex: 6/10) instead of a decimal answer. Your answer does not have to be reduced to the lowest terms (ex: 6/10 is fine instead of 3/5). [2 pts each]

- (a) the syllable sequence *pen/ga*
- (b) the syllable sequence *go/ta*
- (c) the syllable sequence *mor/ly*
- (d) the syllable sequence *go/mor*
- (e) the syllable sequence *kist/ly*

(2) Here is a sequence of syllables Sigmund heard while in the land of the Guins. He gathered enough data that he was able to estimate the transitional probabilities between syllables. Where would a learner who only used transitional probability place word boundaries in the following sequence? To indicate a boundary, list the syllable sequence the boundary would appear between (ex: between “guf” and “fa”). There may be more than one boundary. [2 pts]



(3) Suppose a Guin child had the words below in her proto-lexicon. (Note: A stressed syllable is indicated by a ´ mark over the vowel. Ex: EMphasis = émphasis.)

pentá margón pengó brekún

(a) Would this child be likely to adopt a strategy where words begin with stressed syllables? (Yes or No) [1 pt]

(b) Would this child be likely to adopt a strategy where words end with stressed syllables? (Yes or No) [1 pt]

(c) Given the stress-based strategy the Guin child is likely to use, where would this child put word boundaries in the Guin syllable stream below? You may assume that children who believe words begin with stressed syllables will only put a boundary before the syllable that begins a potential word, while children who believe words end with stressed syllables will only put a boundary after the syllable that ends a potential word. To indicate a boundary, list the syllable sequence the boundary would appear between (ex: between “pen” and “garm”). There may be more than one boundary. [4 pts]

pen gárm merk fet zúrf az zu bérk kíst lo

(d) Given the stress-based strategy the Guin child is likely to use, where would this child be likely to put word boundaries in the *English* syllable stream below? You may assume that children who believe words begin with stressed syllables will only put a boundary before the syllable that begins a potential word, while children who believe words end with stressed syllables will only put a boundary after the syllable that ends a potential word. To indicate a boundary, list the syllable sequence the boundary would appear between (ex: between “guf” and “fa”). There may be more than one boundary. [5 pts]

Sá rah thought that Já reth was a vé ry é le gant mó narch

(4) Sigmund was quite taken with the idea of a computational model of word segmentation and has been testing some different learning strategies out. Here is an input sentence he gave to his model:

“If she ever kisses you, I will turn you into a prince – prince of the land of stench!”

Here are the words his model identified:

Ifshe ever kissesy, Iwill turnyou into aprince – prince oftheland ofstentch!

What is the precision of the model’s performance on this sentence? What is its recall? Your answer should take the form of # / # (ex: 3/4). [4 pts]

Example:

Original: “I like pirates.”

Model’s segmentation: “Ilike pirates.”

Precision calculation: The model identifies 2 words total (*Ilike, pirates*), only one of which is a real word (*pirates*).

Answer: Precision = 1/2.

Recall calculation: The model identifies 1 real word (*pirates*), but should have identified 3 words (*I, like, pirates*).

Answer: Recall = 1/3.

(5) Sigmund has been observing how the Guin use words to refer to things in their environment, and was very interested in the idea that cross-situational learning relies on Bayesian inference. He has been trying to determine the probability that the word “pengo” refers to different objects.

(a) In the first utterance he heard, Sigmund observed that there were three objects present (object1, object2, and object3), and three words were used: “pengo”, “margon”, and “brekun”.

(1) Assuming all word-referent mappings are equally likely before Sigmund has heard anything, what is the **prior probability** that “pengo” refers to object1, given this utterance? What about the prior probability that “pengo” refers to object2? What about the prior probability that “pengo” refers to object3? (Hint: At this point, “pengo” could refer to object1, object2, or object3.) Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [3 pts]

(2) What is the **likelihood** that Sigmund would observe this utterance, if “pengo” referred to object1? What about if “pengo” referred to object2? What about if “pengo” referred to object3? Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [3 pts]

(3) What is the **posterior probability** that “pengo” refers to object1, given this utterance? What about the posterior probability that “pengo” refers to object2, given this utterance? What about the posterior probability that “pengo” refers to object3, given this

utterance? Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [3 pts]

(b) In the second utterance he heard, Sigmund observed that there were two objects present (object2 and object4), and two words were used: “pengo” and “merkel”.

(1) Given that Sigmund has observed the first utterance, what is the **prior probability for the second utterance** that “pengo” refers to object1, given this utterance? What about the prior probability that “pengo” refers to object2? What about the prior probability that “pengo” refers to object3? What about the prior probability that “pengo” refers to object4? (Hint: At this point, “pengo” could refer to object1, object2, object3, or object4.) Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [4 pts]

(2) What is the **likelihood** that Sigmund would observe this utterance and the previous utterance, if “pengo” referred to object1? What about if “pengo” referred to object2? What about if “pengo” referred to object3? What about if “pengo” referred to object4? Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [4 pts]

(3) What is the **posterior probability** that “pengo” refers to object1, given this utterance and the previous utterance? What about the posterior probability that “pengo” refers to object2? What about the posterior probability that “pengo” refers to object3? What about the posterior probability that “pengo” refers to object4? Give each of your answers in the form of a fraction if it is not 0 or 1. Ex: 1/10. [4 pts]

(6) Sigmund has used a frequent frames strategy on some Guin data, and identified the following clusters of words:

Cluster 1	Cluster 2	Cluster 3
pengo	gar	pengamut
gar	tirum	zuma
firog	firog	berno
berno	pengamut	gar

He asked his Guin friends and found out that the words in group 1 below are nouns while the words in group 2 below are verbs:

Group 1 (Nouns)	Group 2 (Verbs)
pengo	berno
zuma	pengamut
firog	tirum
gar	
zu	

Example calculation: What are the precision and recall for cluster 2 if Sigmund compares it against the Group 2 Verbs category?

Precision calculation:

Cluster 2 identifies 4 words, 2 of which are in the Group 2 Verbs category (*tirum*, *pengamut*).

Answer: Precision = 2/4.

Recall calculation:

Cluster 2 identifies 2 of the words that are in the Group 2 Verbs category (*tirum*, *pengamut*), and the Group 2 Verbs category has 3 words in it.

Answer: Recall = 2/3.

For all precision and recall calculations below, your answer should take the form of # / # (ex: 3/4).

- (a) What are precision and recall for cluster 3 if Sigmund compares it against the Group 1 Nouns category? [4 pts]
- (b) What are the precision and recall for cluster 3 if Sigmund compares it against the Group 2 Verbs category? [4 pts]
- (c) Suppose that two clusters must share 50% or more of their members in order to be merged. Should Cluster 1 be merged with Cluster 2, given this criterion? (Yes or No) [1 pt]
- (d) Suppose Cluster 1 is merged with Cluster 2. What are the precision and recall for this new merged cluster (Cluster 1 + Cluster 2) if Sigmund compares it against the Group 2 Verbs category? [4 pts]
- (e) Suppose Cluster 1 is merged with Cluster 2. What are the precision and recall for this new merged cluster (Cluster 1 + Cluster 2) if Sigmund compares it against the Group 1 Nouns category? [4 pts]