

Psych156A/ Ling150

Spring 2010

Homework 2: Words and Morphology

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.

(33 points total)

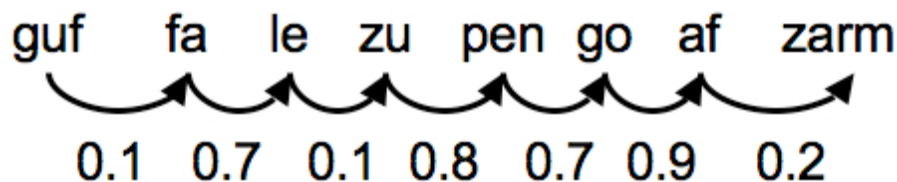
(1) Sigmund has brought back some information about the language of Guin. Here are all the words he found that begin with “pen”, separated by syllables (there is a / between syllables):

pen/tam	=	“pentam”
pen/go/mor	=	“pengomor”
pen/ta/kist/ne	=	“pentakistne”
pen/ma/ly	=	“penmaly”
pen/go/kist/ly	=	“pengokistly”
pen/go/ta/kist/ne	=	“pengotakistne”
pen/go	=	“pengo”
pen/tam/kist/ly	=	“pentamkistly”

Calculate the transitional probability of the syllable sequences below, given these data. You should write your answer as # / # (ex: 6/8) instead of a decimal answer. [1 pt each]

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

(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) 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:

“You have to understand my position. I am a coward, and Jareth scares me.”

Here are the words his model identified:

Youhaveto understand my position. Iama coward, and Jareth scaresme.”

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.

(4) 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
zuma	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	gar
	tirum
	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, 3 of which are in the Group 2 Verbs category (*gar, tirum, pengamut*).

Answer: Precision = 3/4.

Recall calculation:

Cluster 2 identifies 3 of the words that are in the Group 2 Verbs category (*gar, tirum, pengamut*), and the Group 2 Verbs category has 5 words in it.

Answer: Recall = 3/5.

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 1 if Sigmund compares it against the Group 1 Nouns category? [4 pts]
- (b) What are the precision and recall for cluster 1 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 3, given this criterion? [1 pt]
- (d) Suppose Cluster 1 is merged with Cluster 3. What are the precision and recall for this new merged cluster (Cluster 1 + Cluster 3) if Sigmund compares it against the Group 2 Verbs category? [4 pts]
- (e) Suppose Cluster 1 is merged with Cluster 3. What are the precision and recall for this new merged cluster (Cluster 1 + Cluster 3) if Sigmund compares it against the Group 1 Nouns category? [4 pts]

(5) Sigmund was interested in the idea of irregular rules, and is trying to see if Guin children show evidence of having irregular rules. He has gathered the following data:

Potential irregular rule 1: *ero → *erotem

of verbs following this rule: 20

examples of verbs following this rule:

pengero → pengerotem

trikero → trikerotem

vlero → vlerotem

of times Guin children have heard any verbs following this rule: 2000

of times Guin children have heard “trikerotem”: 10

Potential irregular rule 2: *vor → *var

of verbs following this rule: 22

examples of verbs following this rule:

movor → mover

plovor → plovar

trikovor → trikovar

of times Guin children have heard any verbs following this rule: 200

of times Guin children have heard “trikovar”: 10

(a) If children do not have these irregular rules in their minds, would you expect their performance on “trikerotem” and “trikovar” to be about the same, or would you expect it to be different? Briefly explain why. [2 pts]

(b) If children have these irregular rules in their minds, would you expect their performance on “trikerotem” and “trikovar” to be about the same, or would you expect it to be different? Briefly explain why. [2 pts]

(c) Sigmund gathered some data on how well children produce verbs that follow these potential irregular rules:

Potential irregular rule 1: *ero → *erotem

% of the time Guin children produce “trikoretm” correctly: 75%

Potential irregular rule 2: *vor → *var

% of the time Guin children produce “trikovar” correctly: 25%

Do these data support the existence of irregular rules for the past tense in children’s minds? Briefly explain why. (Hint: Think about what you answered in parts (a) and (b).)

[2 pts]