

Psych156A/ Ling150

Winter 2009

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.

(58 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”:

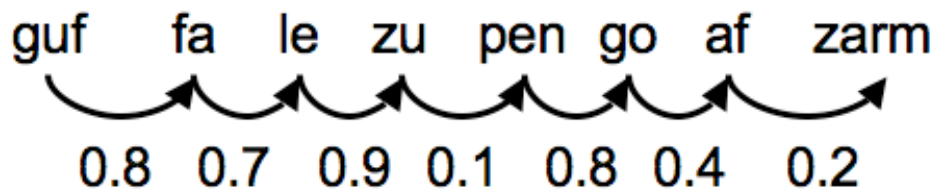
pengo pentumbry pentanorla pentakist

pentanor pentanorne penlig pencrom

Calculate the transitional probability of the words or syllable sequences below, given these data. You should feel free to write your answer as # / # (ex: 6/8) instead of a decimal answer. [2 pts each]

- (a) the word “pencrom”
- (b) the syllable sequence “penta”
- (c) the syllable sequence “tanor”
- (d) the syllable sequence “penmo”

(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:

“For my will is as strong as yours and my kingdom is as great.”

Here are the words his model identified:

For mywill isas strongasyours and mykingdom isasgreat.

What is the precision of the model’s performance on this sentence? What is its recall? Make sure to show how you calculated these measures, based on the words in the original sentence. Your answer should take the following form:

Precision = # / # (ex: 4/8)

Explanation for each number. (ex: 4 was in the numerator because...8 was in the denominator because...)

Recall = # / # (ex: 4/8)

Explanation for each number. (ex: 4 was in the numerator because...8 was in the denominator because...)

[8 pts]

(4) Sigmund has just heard about the idea of algebraic learning for word segmentation, and wants to see how it would work on some of his Guin data.

Here are some Guin words he knows:

pengo zu pencrom pentanor az

(a) Where would an algebraic learner put word boundaries in the syllable sequence below? Your answer should indicate a boundary by identifying the syllable sequence the boundary would come between (ex: between “pen” and “go”). There may be more than one boundary. [3 pts]

Sequence: pen go az la to pen crom mer tem pen ta nor

(b) Sigmund then read Gambell & Yang (2006) and saw their description of the Unique Stress Constraint for word segmentation. He wants to try it out on the same sequence of Guin data. This time, however, he has stress information.

Here are the Guin words Sigmund knows, along with their stress patterns:

(Note: pén = the “pen” syllable has stress)

péngo zú péncrom péntanor áz

Where would an algebraic learner with knowledge of the Unique Stress Constraint put word boundaries in the syllable sequence below? Your answer should indicate a boundary by identifying the syllable sequence the boundary would come between (ex: between “pen” and “go”). There may be more than one boundary. [3 pts]

Sequence: pén go áz lá tó pén crom mér tem pén ta nor

(5) Sigmund liked the idea that a word’s category (like noun or verb) is really just a description of the way that word can be used in the language. Given the following contexts, tell Sigmund whether you think the novel word in each example (indicated by CAPITALS) is a noun (like “goblin”), a verb (like “sing”), an adjective (like “hot”), or an adverb (like “dreamily”). Be sure to briefly explain why you think so. (You might find it helpful to substitute words you know in place of the novel words, and see which ones fit best.) [2 pts each]

- (a) That’s a very BOFT bog.
- (b) Is Sir Didymus REKKing the Bog of Eternal Stench?
- (c) Ludo shouldn’t have STROOPed his paw in the bog.
- (d) They had to beware the BREER’s tricks.

(6) Sigmund has once again been observing the Guins, and has been trying to identify which words behave similarly on the basis of the frequent frames he has encountered. Here an example utterance that Sigmund has from the Guin language.

“felgo bofty mu az berg mu merk berg felgo zu mu var berg az porto mu freggo berg felgo seech mu set berg draz pino trem felgo trem mu peri berg lootem bleeter.”

- (a) What is the most frequent frame in this utterance? [2 pts]
- (b) Which words would this frame cluster together? [2 pts]
- (c) What is the second most frequent frame in this utterance? [2 pts]
- (d) What words would this second most frequent frame cluster together? [2 pts]

(7) By using the frequent frames learning strategy on a large set of Guin data, Sigmund was able to identify 120 words that seem to behave like nouns. When he asked his Guin friends about them, he found out that 65 of the 120 words he identified did belong to the same category (nouns). However, from the data set he had, he should have found 35 more nouns than he actually did correctly identify. Calculate the precision and recall for Sigmund’s frequent frames approach on that set of Guin data. Be sure to show how you calculated each one. Your answer should take the following form:

Precision = # / # (ex: 4/8)

Explanation for each number. (ex: 4 was in the numerator because...8 was in the denominator because...)

Recall = # /# (ex: 4/8)

Explanation for each number. (ex: 4 was in the numerator because....8 was in the denominator because...)

[8 pts]

(8) Sigmund has been trying to figure out yet more of his Guin data. He's been particularly interested in how the past tense is formed. Here is the data:

root form	past tense form
pengar = "to slip"	merpengar = "slipped"
dappler = "to sing"	merdappler = "sang"
broll = "to stumble"	merbroll = "stumbled"
pooket = "to grin"	merpooket = "grinned"

(a) Given only these data, how is the past tense formed in Guin? (That is, what do you have to do to a verb to make it into its past tense form?) [1 pt]

(b) Now, consider these additional data Sigmund discovered:

tokk = "to touch"	tokkem = "touched"
sokk = "to drop"	sokkem = "dropped"
vokk = "to throw"	vokkem = "threw"

Do these follow the rule you described in part (a)? [1 pt]

(c) Suppose a Guin child encountered only the data from part (b). What rule might this child hypothesize for the past tense? [1 pt]

(d) Is the new rule you described in part (c) compatible with the data from part (a)? [1 pt]

(Bonus) (+2 pts if correct): Suppose a Guin child encountered the data in both part (a) and the data in part (b). How could a child make sense of these data? (Hint: Think about the regular and irregular past tense forms in English, and what determines if a verb is regular or irregular. Which set of verbs above is likely to be an irregular verb class in Guin?)

(9) 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: *irk → *ark

of verbs following this rule: 15

examples of verbs following this rule: vlirk~vlark, hzirk~hzark, trirk~trark

of times Guin children have heard any verbs following this rule: 2500

of times Guin children have heard "hzark" : 10

Potential irregular rule 2: *gor → *gorm

of verbs following this rule: 18

examples of verbs following this rule: mogor~mogorm, bargor~bargorm, ...

of times Guin children have heard any verbs following this rule: 100

of times Guin children have heard “mogorm”: 10

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

(b) If children have these irregular rules in their minds, would you expect their performance on “hzark” and “mogorm” to be about the same, or would you expect it to be different? Why? [2 pts]

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

Potential irregular rule 1: *irk → *ark

% of the time Guin children produce “hzark” correctly: 80%

Potential irregular rule 2: *gor → *gorm

% of the time Guin children produce “mogorm” correctly: 20%

Do these data support the existence of irregular rules for the past tense in children’s minds? (Hint: Think about what you answered in parts (a) and (b).) [2 pts]