Psych 156A/ Ling 150: Psychology of Language Learning

Lecture 12

Learning Biases

Announcements

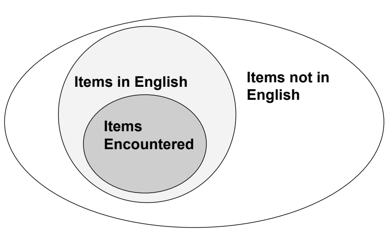
Be working on HW3 (due: 2/24)

Pick up previous HWs and midterm if you haven't done so

Reminder: Sean's office hours this week are on Friday, 10am - 12pm

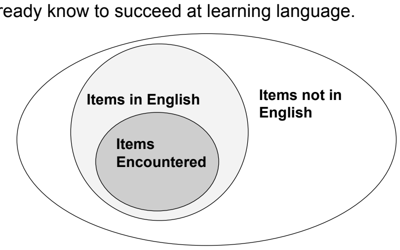
Summary from last time: Poverty of the Stimulus and Learning Strategies

Poverty of the stimulus: Children will often be faced with multiple generalizations that are compatible with the language data they encounter. In order to learn their native language, they must choose the correct generalizations.



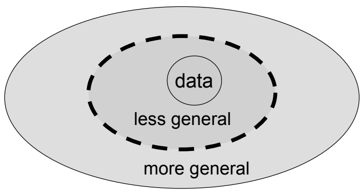
Summary from last time: Poverty of the Stimulus and Learning Strategies

Claim of prior (innate) knowledge: Children only seem to make the right generalization. This suggests *something* biases them to make that generalization over other possible generalizations. Importantly, that *something* isn't available in the data itself. It is knowledge they must already know to succeed at learning language.



Summary from last time: Poverty of the Stimulus and Learning Strategies

One Learning Bias: Experimental research on artificial languages (Gerken 2006) suggests that children prefer the more conservative generalization compatible with the data they encounter.



Learning Biases

"Innate capacities may take the form of biases or sensitivities toward particular types of information inherent in environmental events such as language, rather than a priori knowledge of grammar itself." - Seidenberg (1997)

Example: Children seem able to calculate transitional probabilities across syllables (Saffran, Aslin, & Newport 1996).

Example: Adults seem able to calculate transitional probabilities across grammatical categories (Thompson & Newport 2007)

But is it always just statistical information of some kind?

Gambell & Yang (2006) found that tracking transitional probabilities across syllables yields very poor word segmentation on realistic English data.

Other learning strategies like the Unique Stress Constraint and algebraic learning did far better. These other learning strategies were not statistical in nature - they did not use probabilistic information available in the data.

Peña et al. 2002: Experimental Study

Goal: examine the relation between statistical learning mechanisms and non-statistical learning mechanisms (like algebraic learning).

Adult learners' task on artificial language:

(1) word segmentation

(2) generalization about words in the language (somewhat similar to categorization)

PURAKIBELIGATAFODUPUFOKITALIDUBERAGA...

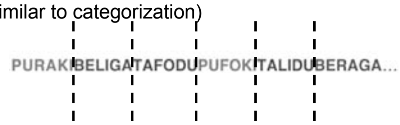
Peña et al. 2002: Experimental Study

Goal: examine the relation between statistical learning mechanisms and non-statistical learning mechanisms (like algebraic learning).

Adult learners' task on artificial language:

(1) word segmentation

(2) generalization about words in the language (somewhat similar to categorization)



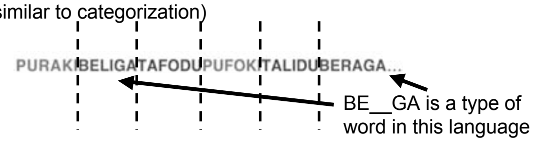
Peña et al. 2002: Experimental Study

Goal: examine the relation between statistical learning mechanisms and non-statistical learning mechanisms (like algebraic learning).

Adult learners' task on artificial language:

(1) word segmentation

(2) generalization about words in the language (somewhat similar to categorization)



Peña et al. 2002: Experimental Study

The artificial language: "AXC language"

Syllables: A, X, C

Generalization:
A perfectly predicts C: A_C is a word in the language
pu_ki, be_ga, ta_du

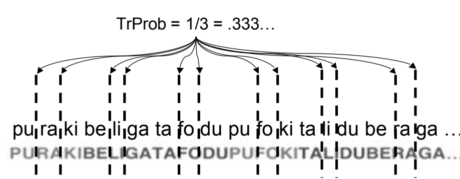
Intervening syllable X: _ra_, _li_, _fo_

pu ra ki be li ga ta fo du pu fo ki ta li du be ra ga …
PURAKIBELIGATAFODUPUFOKITALIDUBERAGA…

Peña et al. 2002: Experimental Study

The artificial language: "AXC language"

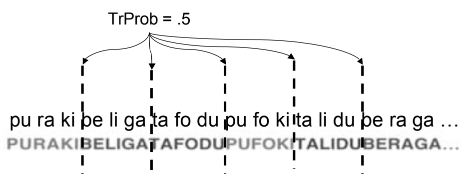
Note: transitional probability information is not informative.



Peña et al. 2002: Experimental Study

The artificial language: “AXC language”

Note: transitional probability information is not informative.

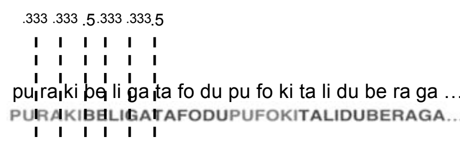


Peña et al. 2002: Experimental Study

The artificial language: “AXC language”

Note: transitional probability information is not informative.

TrProb is actually higher at word boundaries…

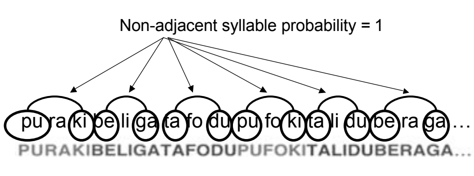


Peña et al. 2002: Experimental Study

The artificial language: “AXC language”

Note: transitional probability information is not informative.

Only non-adjacent syllables are informative about what words are in the language.



First Question: Good word segmentation?

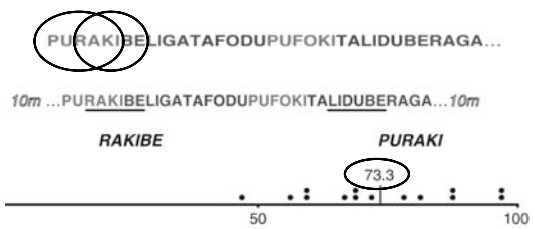
10 minute familiarization period

Can adults recognize words from part-words?
Remember: transitional probability won’t help - it’ll bias them the wrong way.

word: pu ra ki
TrProb (pura) = 1/3, TrProb (raki) = 1/3,
TrProb (puraki) = TrProb(pura)*TrProb(raki) = 1/3*1/3 = 1/9

part-word: ra ki be
TrProb(raki) = 1/3, TrProb (kibe) = 1/2
TrProb (rakibe) = TrProb(raki)*TrProb(kibe) = 1/3*1/2 = 1/6 (higher than 1/9)

First Question: Good word segmentation?



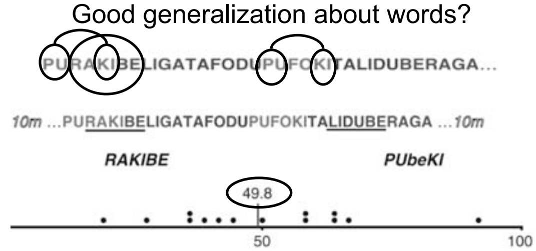
Adults prefer real words to part-words that they actually heard. This means they can unconsciously track the non-adjacent probabilities of the AXC language and identify the words.

Next Question: Good generalization about words?

Do adults make the generalization about which words belong in the language (ex: that words of the language can take the form of pu_ki)

Compare:	rakibe	vs.	pubeki
	Part-word heard during training		Word that follows the pattern, but is not a word heard during training

Next Question: Good generalization about words?



Adults have no preference between part-words that they actually heard and real words that follow the generalization about words in the language, but which they didn't actually hear. This means they can't use the non-adjacent probabilities of the AXC language to identify properties of the words in general.

What's going on?

"We conjecture that this reflects the fact that the discovery of components of a stream and the discovery of structural regularities require different sorts of computations…the process of projecting generalizations…may not be statistical in nature." - Peña et al. (2002)

Prediction for Different Types of Computation

"…it is the type of signal being processed rather than the amount of familiarization that determines the type of computation in which participants will engage…changing a signal even slightly may induce a change in computation." - Peña et al. (2002)

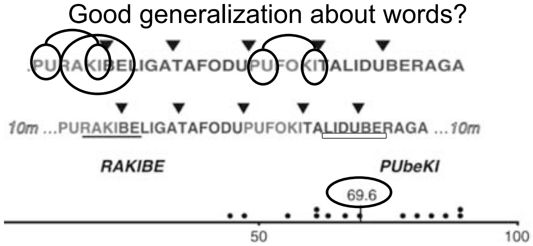
Types of computation: statistical, algebraic

New Stimuli: Stimulating Algebraic Computation?

10 minute familiarization period with 25ms (subliminal) gaps after each word

If word segmentation is already accomplished, subjects will be free to engage their algebraic computation. This should allow them to succeed at identifying the properties of words in the artificial language (e.g. pu_ki, be_ga, ta_du), since this kind of structural regularity is hypothesized to be found by algebraic computation.

Question: Good generalization about words?



Adults prefer real words that follow the generalization about words in the language, but which they didn't actually hear, over part-words they did hear. This means they can use the non-adjacent probabilities of the AXC language to identify properties of the words in general. They make the structural generalization.

Prediction: Algebraic vs. Statistical

Idea: Subjects are really using a different kind of computation (algebraic) because of the nature of the input. Specifically, the input is already subliminally segmented for them, so they don't need to engage their statistical computation abilities to accomplish that. Instead, they are free to (unconsciously) notice more abstract properties via algebraic computation.

Prediction 1: If the words are *not* segmented subliminally, statistical computation will be invoked. <u>It doesn't matter if subjects hear a lot more data</u>. Their performance on preferring a real word they didn't hear over a part-word they did hear will not improve.

Question:
Good generalization about words?

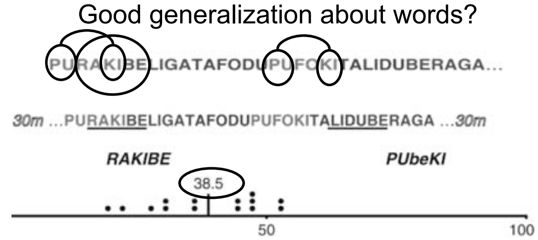
If given 30 minutes of training on unsegmented artificial language, do adults fail to make the generalization even though they have a lot more data?

Compare: rakibe vs. pubeki

rakibe: Part-word heard during training a lot!

pubeki: Word that follows the pattern, but is not a word heard during training

Question:
Good generalization about words?



If given 30 minutes of training on unsegmented artificial language, adults really prefer part-words that they actually heard over real words that follow the generalization about words in the language, but which they didn't actually hear. They can't make the generalization: prediction 1 seems true.

Prediction: Algebraic vs. Statistical

Idea: Subjects are really using a different kind of computation (algebraic) because of the nature of the input. Specifically, the input is already subliminally segmented for them, so they don't need to engage their statistical computation abilities to accomplish that. Instead, they are free to notice more abstract properties via algebraic computation.

Prediction 2: If the words *are* segmented subliminally, algebraic computation will be invoked. <u>It doesn't matter if subjects hear a lot less data</u>. They will still prefer a real word they didn't hear over a part-word they did hear.

Question:
Good generalization about words?

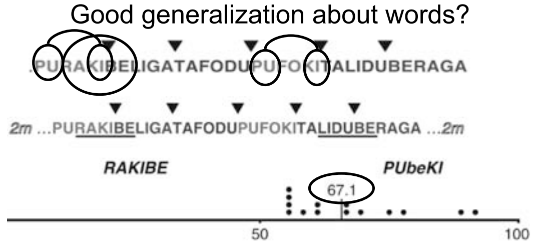
If given 2 minutes of training on segmented artificial language, do adults make the generalization even though they have a lot less data?

Compare: rakibe vs. pubeki

rakibe: Part-word heard during training

pubeki: Word that follows the pattern, but is not a word heard during training

Question:
Good generalization about words?



If given 2 minutes of training on segmented artificial language, adults really prefer real words that follow the generalization about words in the language, but which they didn't actually hear, over part-words that they actually heard. They still make the generalization: prediction 2 seems true.

Peña et al. (2002): Summary

While humans may be able to compute powerful statistical relationships among the language data they're exposed to, this may not be enough to capture all the linguistic knowledge humans come to possess.

In particular, learning structural regularities (like structural properties of words) may require a non-statistical learning mechanism, perhaps algebraic computation.

Different kinds of computation can be cued in learners based on the data at hand - this is a learning bias (to use a particular computation given particular data). Statistical computation was cued by the need to group and cluster items together. Algebraic computation was cued once items were already identified, and generalizations had to be made among the items.

What kind of things can statistical computation keep track of?

Idea: "Learners might be able to compute certain types of statistical regularities, but not others." - Newport & Aslin (2004)

Important: AXC-syllable language (statistical regularity between 1st and 3rd syllable of the word, like what Peña et al. 2002 used) does not naturally occur in real languages.

What kind of non-adjacent regularities do real languages actually exhibit? Can humans reliably segment these kinds of languages using statistical computation?

Naturally occurring non-adjacent regularities

Example of non-adjacent dependency: between individual segments (sounds)

Semitic languages: words built from consonantal "stems", where vowels are inserted to make different words

Arabic: k-t-b = "write"

kataba = "he wrote" yaktubu = "he writes"

kitaab = "book" maktab = "office"

Non-adjacent segment regularities: consonants

Newport & Aslin (2004): AXCXEX segment language
p_g_t, d_k_b filler vowels in IPA: a, i, æ, o, u, e
(generalization about words)

Subject exposure time to artificial language made up of these kinds of words: 20 minutes

Result 1: Subjects were able to segment words based on non-adjacent segment regularities. This is similar to the result found in Peña et al. 2002 on their artificial language.

Non-adjacent segment regularities: vowels

Newport & Aslin (2004): XBXDXF segment language
_a_u_e, _o_i_æ filler consonants: p, g, t, d, k, b
(generalization about words)

Subject exposure time to artificial language made up of these kinds of words: 20 minutes

Result 2: Subjects were again able to segment words based on non-adjacent segment regularities. So this again accords with the results found by Peña et al. (2002).

Newport & Aslin (2004): Summary

When subjects are tested with artificial languages that reflect properties real languages have (such as statistical dependencies between non-adjacent segments), they are still able to track statistical regularities.

This suggests that statistical computation is likely to be something real people use to notice the statistical regularities (non-adjacent or otherwise) that real languages have. It is not just something that will only work for the regularities that have been created in a lab setting, such as those between non-adjacent syllables in artificial languages.

Questions?