Psych 229:
Language Acquisition

Lecture 6
Words & Models

Gambell & Yang 2006: Computational model of word segmentation

What happened?

Gambell & Yang 2006: Computational model of word segmentation

Would more data help?…probably not

Gambell & Yang 2006: Computational model of word segmentation

What about other models (Swingley (2005)) that have success on data like this?

Here’s an idea…(and a language-independent one at that)

Why do they think this might work?

Is this the only language-independent constraint?

How easy is it to identify “main stress” in a word (especially cross-linguistically)?
Gambell & Yang 2006: Computational model of word segmentation

Here’s one model...

Hey, not bad!

The algorithm is agnostic, so the outcome is determined by the input.

What about algebraic learning?

Strong Weak1 Weak2... Weakn Strong

S W1 = known word
W2... Wn = known word

Agnostic: ignore this string
Random: pick a division point at random

Conclusions

Constraints on learning (innate biases)
Gambell & Yang 2006: Computational model of word segmentation

Discussion Questions

What about other languages besides English? (Turkish, Mohawk - polysynthetic languages)

What does it mean that the USC+Algebraic learner actually identifies words much quicker than real children seem to? [Bruno]

Gómez & Lakusta 2004: Categorization

Gómez & Lakusta 2004: Categorization

What babies can do...

What babies might do...

What babies can do...

What babies might do...

Gómez & Lakusta 2004: Categorization

Another Idea: Distributional Learning

One Idea: Semantic Bootstrapping

Category abstraction task

Previous work (aX, bY paradigm)