Recap

- We have so far looked at the words and rules theory.
- Different models of past tense formation.
- Perceptrons and neural networks.
- Back to language and how words emerge in the first place.
- Speech segmentation.

Reading:


The Development of Language

- **Vegetative sounds** (0-6 weeks)
  - Cooing (6 weeks)
  - Laughter (16 weeks)
  - Vocal play (16 weeks-6 months)
  - Babbling (6 months-10 months)

- **Single word utterances** (10-18 months)
  - Two-word utterances speech (18 months)
  - Telegraphic speech (2 years)
  - Full sentences (2 years 6 months)

[https://www.youtube.com/watch?v=YI1aPCdJaMw](https://www.youtube.com/watch?v=YI1aPCdJaMw)

[http://www.youtube.com/watch?v=_JmA2ClUvUY](http://www.youtube.com/watch?v=_JmA2ClUvUY)
**How Do We Learn Words?**

- Knowing a language implies having a **mental lexicon**.
- Memorized set of associations among sound sequences, their meanings, and their syntax.
- Speech stream lacks any acoustic analog of the **blank spaces** between printed words.
- Basic units of linguistic input are not words but **entire utterances**.
- Child's task: to **discover the words** themselves in addition to meaning and syntax.

**What do Infants Hear?**

- Where are you going?
- How does a bunny rabbit walk?
- Does she walk like you or does she go hop hop hop?
- What are you doing?
- Sweep broom.
- Is that a broom?
- Though 'twas a brush.

Adam's mother (Brown, 1973)

**Where Are the Words?**

- THEREDONATEA KETTLE OF TEN CHIPS
- THE RED ON A TEA KETTLE OFTEN CHIPS
- THERE, DONATE A KETTLE OF TEN CHIPS
- THERE, DONATE A KETTLE OF TEN CHIPS

**Important Questions**

- How does an infant divide the input into reusable units?
- How does she represent those units?
- What does she know about them and when?

Not an end in itself: provides useful units (Peters, 1983) for learning a grammar: lexicon, morphosyntax, phonology.
How do Infants Segment Speech?

Infants make use of multiple cues in the input, most popularly:
- **Stress patterns**: English usually stresses 1st syllable, French always the last; final syllables of words are longer (ham*ster* vs. ham).
- **Phonotactic constraints**: every word must contain a vowel, finite set of consonant clusters that can occur at the beginning of a word, before the first vowel (g*dog* is not a possible English word).
- **Statistical regularities**: within words, there is a consistent sequence of elements.
- **Bootstrapping** from known words.

Transitional Probability

Words create regularities in the sound sequences of a language.
- There is a consistent sequence of elements within words
- Sequences that don’t occur within words can only occur at word boundaries.
- Sequences that don’t occur within a word will tend to occur infrequently.
- Thus, we can find word boundaries by looking for unlikely transitions.

Suppose the phoneme [ð] occurs 200,000 times in a text:
- 190,000 times are before a vowel (as in the, this);
- 200 times are before [m].

Transitional Probability

\[
P(\text{y} | \text{x}) = \frac{p(x, y)}{p(x)} \approx \frac{\text{freq}(x, y)}{\text{freq}(x)}
\]
Saffran et al. (1996) asked whether 8-month-old infants can extract information about word boundaries solely on the basis of statistical information.

- Create “language” from nonsense words.
- Infants listen to synthesized language (tokibu, gikoba).
- Then, test: can infants distinguish words (tokibu) vs. part-words (bugikobu)?

Infants are exposed for 2 minutes to nonsense language (tokibu, gopila, gikoba, tipolu).

- Only statistical cues to word boundaries
- Then record how long they attend to novel sets of stimuli that either do or do not share some property with the familiarization data.
- Discrimination between words and part-words (sequences spanning word boundaries)
- If there’s a difference, there has been some learning during familiarization.
Headturn Preference Procedure

Results

- Infants show longer listening times for part-words
- Infants can extract information about sequential statistics of syllables (input contained no pauses, intonational patterns)

Conclusions

- Humans can use statistical information to segment speech.
- But all words were trisyllabic
- So, transitional probabilities were either 1 or .33
- Will this work if these are varied in a more naturalistic way?

Lexicons and Segmentation

- The use of transitional probabilities to do word segmentation ignores the fact that words are being learned at the same time.
- There are statistical methods for speech segmentation that incorporate the learning of a lexicon as a subcomponent.
- Brent and Cartwright (1996): find the lexicon which minimizes the description of the observed data

Minimum Description Length

\[ \text{size(description)} = \text{size(lexicon)} + \text{size(data-encoding)} \]
MDL and Lexicons

Minimum Description Length

\[ \text{size(description)} = \text{size(lexicon)} + \text{size(data-encoding)} \]

- The MDL principle minimizes the length of words
- Shorter words are more plausible
- Minimizes the number of different words
- Try to make use of words you already know
- Maximizes the probability of each word
- Words recur as often as possible

Brent and Cartwright (1996)

Input

doyouseethekitty
seethekitty
doyoulikethekitty

Segmentation

doyou see the kitty
see the kitty
doyou like the kitty

Lexicon

1 do 2 the 3 you
4 like 5 see
6 kitty

Derivation

1 3 5 2
5 2
1 3 4 2

Length: 26+13=39

MDL model is tested on (phonetically) transcribed speech from the CHILDES corpus.

- An idealization of the raw acoustic signal.
- Model searches for segmentation of the input with least MDL.
- Search algorithm is not incremental; it reads in the entire input before segmenting any part of it.
- Approach does not rely on language-specific input!
- Computational simulations systematically explore hypothesis that distributional regularity is useful for word segmentation.
In order to acquire a lexicon young children segment speech into words using multiple sources of support; focused on distributional regularities.

- transitional probability provides cues
- verified by Saffran et al. (1996) experiments
- computational model of word segmentation
- based on Minimum Description Length Principle

Next lecture: word learning.