Semantic Expectations in Child Language Comprehension

Gleb Promokhov
gxp5819@rit.edu

Abstract

As adults people are able to comprehend the correct meaning of noisy speech on the fly. Errors thrown by regular speech and semantic mistakes are caught by the listener's expectations of the utterance's meaning as a whole. Adults have been shown to rely on these top-down semantic expectations as well as bottom-up phonetic cues when comprehending sentences. This experiment explores how children develop this skill, using a probabilistic modeling of their speech comprehension.

1 Introduction

Language comprehension is the process of human perception which recognizes sounds we hear as phonemes, groups them into words, organizes their syntax in sentences and deduces its semantic meaning. The brain is able to do this incredibly quickly thanks in part to incremental comprehension, where we actively predict the content of an utterance as we hear it. Not only does this speed up comprehension, but allows people to resolve noise. Given that enough of an utterance was clearly comprehended, corrupted parts of that utterance could be substituted with a listener's guess as to what the original, clear signal was.

Spoken communication is an inherently noisy task, people say the wrong thing, stutter, flip around syllables, mispronounce words, etc... This erroneous speech, amongst the hundreds of other sounds we hear, requires that this noise is somehow filtered out to correctly understand what is being said. The two popular comprehension models, garden path models (Frazier, 1978) and constrain based models (MacDonald, Pearlmutter and Seidenberg, 1994), could have been used for this experiment but fail to capture this concept of noise effectively. The noisy channel model (Shannon, 1948) provides a solution with a model which takes a signal and convolves it with a source of noise to output a corrupted signal, and relates these two signals with a basic Bayesian proportion. In terms of natural language these signals are spoken utterances. Comprehension of noisy speech using this model can be formulated as a conditional probability where given a corrupted utterance *s* you find the likelihood of the matching correct utterance *t*

$$P(t|s) \propto P(s|t)P(t) \tag{1}$$

Where *P(t)* is the listener's language model and $P(s|t)$ is a noise model which captures speech production errors. The noise model can be broken down further into $P(s_p|t) * P(s_s|t)$ where $P(s_p|t)$ is the listener's phonetic interpretation $P(s_s|t)$ is the semantic expectation of the corrupted utterance. In this experiment children will be presented with a eye-tracking picture task, with control and test trials that introduce noise and measure expectation.

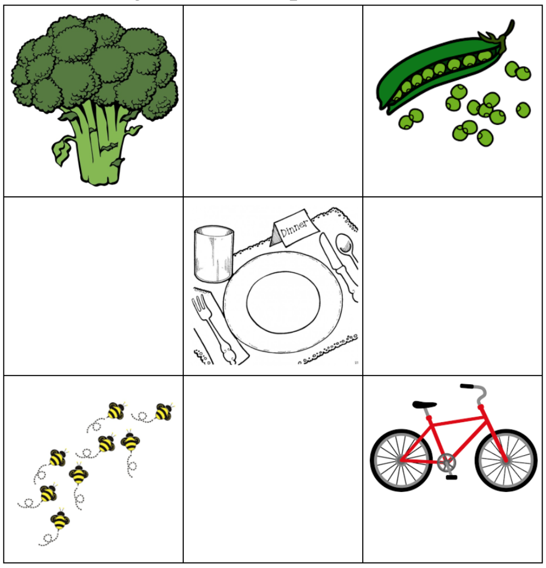
2 Prior Work

Extending this simple proportion to our complex prediction abilities needs certain assumptions resolved. Modeling the phonetic and semantic cues independently can be done with prior work demonstrating how children sometimes fail to combine these cues in eye-tracking experiments when expecting utterances from certain speakers (Graham, Sedivy and Khu, 2013) (Creel, 2012). Children seem to rely on a expected consistency of meaning with speaking partners and are confused when the context signals they were expose to don't match what they hear. This discrepancy shows a separation of the phonetic and semantic processing.

Expectations around noisy utterances has been effectively measured in adults in an on-line questionnaire, tallying answers to comprehension questions about a short phrase between control and test trails with selected syntax errors. (Gibson, Bergen and Piantadosi, 2008). Even with being able to re-read the phrase as much as needed subjects still incorrectly answered questions to clear and corrupted phrases.

Researchers were able to expand on previous grammar development metrics by using noisy

Figure 1: Example Picture Task



channels models to measure the divergence between children's and an model's speech (Sahakian and Snyder, 2013). Errorful children's speech was collected and certain utterances were paired with an corrected for syntax mistakes 'adult' version through crowd-sourced translations. As children age the noisy channel between the speech pair is narrowed.

A separate experiment showed that children will tend to rely on semantic cues of noisy speech when distinguishing phonetically near terms (Yurovksy, Case and Frank, 2017). Adults are much better able to comprehend around ambiguity and static noise than children are, as effectively modeled with a similar Bayesian proportion as this experiment.

3 Experiment

3.1 Subjects

An ideal set of subjects would all have the same native language and be equally distributed between gender, parent's educations, etc... to capture a wide distribution of different stages of language development. Subjects would be grouped together by age to see how their semantic expectations develop as children grow up. For this experiment children between 4 and 10 years of age is ideal, since any younger children have difficultly focusing their gaze making it difficult to make measurements. Any older and children become too effective at resolving noisy ambiguities, reducing the measurement's variance to where differences

Table 1: Term Tallies

	test	control
s_s	$P(s_s \cap t)$	$P(s_s \cap t)$
s_p	$P(s_p \cap t)$	$P(s_p \cap t)$

$$
\begin{aligned}
P(s_p|t) &= P(s_p \cap t)/P(t) \\
P(s_s|t) &= P(s_s \cap t)/P(t) \\
P(s|t) &= P(s_p|t) * P(s_s|t) \\
P(t|s) &\propto P(s|t)P(t)
\end{aligned} \tag{2}
$$

worth exploring are minimal.

3.2 Method

Subjects will be placed in a controlled environment and asked to listen to a statement while looking at a card of five related images, then answer to simple comprehension question. In control trials subjects will listen to a clear, uncorrupted statement. During test trials they will hear a statement replaced with a ambiguous phonetically-near term. This utterance will also be convolved with static noise to accent the error detection[1] and focus the children's attention.

An example includes Figure 1., an image card for the statement: "For dinner I had carrots and _", where during control trials subject hears "peas" and during test trials hears "bees" with static noise. These are shown on the image card along with a semantically related "broccoli" and an unrelated image as controls. Subjects will answer the question "What did I have for dinner?".

3.3 Analysis

The subject's gaze will be tracked and the cumulative time looking at each image will be measured. The image with the highest time will represent their semantic expectation (s_s). Their reply to the question will measure their phonetic comprehension of the statement (s_p). When these measurements match the correct answer it will be tallied into probabilities onto Table 1. Equation 2. shows the basic maths to compute the original $P(t|s)$.

4 Discussion

As kids grow up and live noisier lives, we are exposed to more sounds, used a lot more language

[1] As done in the (Yurovksy, Case and Frank, 2017) study.

and corrected errors along the way. Their ability to correctly expect around ambiguity should increase. This analysis only tallies when the attributes matched the true utterance and will require actual data to explore all the combinations of error possible in the model. It is possible that certain children are better listeners but not guessers than others.

While semantic and phonetic features can be naively assumed as independent, it is still unknown if we truly have separate parallel processes for both when comprehending speech. This experiment predicts to be able to find patterns in phrase long utterances, which could include multiple nouns and references to establish an expected semantic context. There must be some minimal amount of speech is needed to be heard for us be to able to measure its effects. Being able to measure this process at a higher 'resolution', such as being able to measure expectation activations word by word, would lead to more refined results. Expectations of what is being heard are built as the phrase is being heard. Finding this minimal distance would also contribute evidence to the serial/parallel comprehension model debate.

Adults are seemingly able to rely entirely on their expectations of the world around them. A further anecdote to explore are how people sometimes simply 'hear what they want to hear', as in could the brain really match desired phonetic signals with what they actually perceive? Introducing personality into semantic expectations would be very complex to probabilistically measure and model.

People use all of these abilities of comprehension in context, combining phonetic and sensory cues, matching people and places with what they expect to hear and see. In an controlled environment these cues are isolated to an image card and a single utterance. Recognizing patterns in smaller scale comprehension tasks will contribute to understanding perception as a whole.

References

Frazier, L. (1978) On Comprehending Sentences: Syntactic Parsing Strategies Doctoral dissertation *University of Connecticut.*.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994) The lexical nature of syntactic ambiguity resolution. *Psychological review*. 101(4), 676.

Shannon, C. E. (1938) A mathematical theory of communication. *Bell system technical journal*. 27(3), 379-423.

Creel, S. C. (2012) Preschoolers Use of Talker Information in On-Line Comprehension. *Child Development*. 83, 2042-2056.

Yurovsky, D., Case S., and Frank M. C. 2017 Preschoolers Flexibly Adapt to Linguistic Input in a Noisy Channel. *Psychological Science*. 28(1), 132-140.

Gibson, E. and Bergen, L. and Piantadosi, S. T. (2008) Rational integration of noisy evidence and prior semantic expectations in sentence interpretation *National Academy of Sciences*. 110(20), 8051-8056.

Graham, S. A., Sedivy, J., and Khu, M. (2013) That's not what you said earlier: preschoolers expect partners to be referentially consistent. *Journal of Child Language*. 41(1), 34-50.

Sahakian, S., and Snyder, B. (2013) Modeling child divergences from adult grammar. *Transactions of the Association for Computational Linguistics*. 1, 125-138.