Current issues in sentence comprehension: Readings 02, 03 Traxler review article, Levy and Keller 2013

Shravan Vasishth
vasishth@uni-potsdam.de

November 11, 2015

The paper

This paper reviews recent work on anticipation, and good-enough processing theories, focusing on

- surprisal models
- noisy-channel models as an instance of good-enough parsing

I have a very different view about the state of the art than Traxler.

Shravan Vasishth
vasishth@uni-potsdam.de

Classical approaches

From the abstract:
"Classical approaches to parsing describe it [parsing] entirely as a bottom-up signal analysis."
It's true that in psycholinguistics there has been a focus on bottom-up processes until recently. But entirely? How about the work on predictability by Kutas and other researchers using EEGs? I put 13 papers here that document a steady interest in predictability in the EEG world:
http://www.ling.uni-potsdam.de/∼vasishth/courses/Papers.zip.
(password protected)
Note that Levy 2008 does acknowledge this early work by Kutas and others, but it is clearly underappreciated in non EEG psycholinguistics.

Shravan Vasishth
vasishth@uni-potsdam.de

Classical approaches

The Traxler review basically ignores all this rich prior work in the ERP literature, so it's quite a distortion of history to say: "Classical approaches to parsing describe it [parsing] entirely as a bottom-up signal analysis."
He does cite Lau et al 2013 (citation 31), and De Long et al (32) on EEG and prediction, but there was much other relevant work that preceded this that is relevant to prediction.

Shravan Vasishth
vasishth@uni-potsdam.de

Classical approaches

"The idea that syntactic parsing processes can be affected by the probability or likelihood of particular syntactic structures has, in fact, been around for a long time [2,3,42]. Trueswell and colleagues were among the first to provide evidence that the conditional probability of a structural analysis in a given context influenced the processing load imposed on the comprehender [10,43]."
Dan Jurafsky's 1996 paper (A probabilistic model of lexical and syntactic access and disambiguation) is a very important one that is not mentioned in this review.

Shravan Vasishth
vasishth@uni-potsdam.de

Classical approaches

Even the classical paper by Kimball 1973 has as his first principle: "Parsing in natural language proceeds according to a top-down algorithm"
Nevertheless, it is true that predictability and expectation-based processing has recaptured the attention of psycholinguists recently.
As Oscar Wilde remarked (Quoting from memory, so paraphrasing): Everything that can be thought of has been thought of before. All we have now is wonderful adjectives.

Shravan Vasishth
vasishth@uni-potsdam.de

Open questions according to Traxler

"...we need accounts that can tell us how predictions are made, which in turn will tell us why some predictions are made but not others."

- This review missed a big opportunity to point out that the focus has been on predictions, but it is possible that the important thing going on here may be inhibition.
- Some of the EEG articles I provided actually discuss this point and investigate it experimentally.
- It is possible that the "surprise" we see in unexpected continuations is due to the parser trying to inhibit the expected continuation—-if this is true, we should see effects of cognitive control on predictable vs unpredictable continuations (see Bruno Nicenboim's work).

Maybe it's both the greater informativity of the rare event as well as the extra effort at inhibitng the more expected event—this is a wide open area for research.

7

Shravan Vasishth
vasishth@uni-potsdam.de

Three claims regarding prediction (acc. to Traxler)

- First, comprehenders do not have a veridical internal representation of the input [Levy's Cognition paper and the Noise-Channel paper 48,49]. *The role of task seems underappreciated here, and also of capacity differences. See Bruno Nicenboim's work.*
- Second, comprehenders use all available information to compute the likelihood of different interpretations given the cues available in the input [21,24,50?54]. *Seems like an overly confident statement given the evidence out there. I want to believe this though, on commonsense grounds.*
- Third, failed predictions are an important factor in the on-line response to input and in the way that knowledge about language is obtained [50]. *The reference is Hale 2001; I'm not sure what the author means here about knowledge of language being obtained.*

Shravan Vasishth vasishth@uni-potsdam.de

The "noisy channel" model

Traxler introduces the "noisy-channel" model . The basic idea is simply a restatement of Bayes Theorem, so I am not sure why we need a new term for this.
Interpretation $\propto$ Prior Belief $\times$ Data
A simple example is: if you smell some other woman's perfume on your husband's shirt (data), your conclusion about what that mean (interpretation) depends on your prior belief about your husband (consider: newly married vs 25 years of stable marriage; other knowledge about how likely your husband is to be a philanderer).

- Strong priors will override data and dominate the interpretation.
- Lots of data will eventually override even a strong prior, and will eventually dominate interpretation.

Shravan Vasishth
vasishth@uni-potsdam.de

The "noisy channel" model

If the signal is noisy (for internal or external reasons), then the reader will use his prior
Examples:

- *The mother gave the candle the girl* is "less disruptive" than *The mother gave the girl to the candle*. Reason: one can insert a missing *to* in the first sentence: *The mother gave the candle to the girl*.
  Both sentence sentences are anomalous, but subjects can recover some sensible meaning in one but not the other.
  Clearly there is some recovery process happening in one.
- Similarity-based interference is also subsumed by Traxler into the noisy channel model.

Shravan Vasishth
vasishth@uni-potsdam.de

Traxler's critique of the "noisy channel" model

See his p. 4.

Shravan Vasishth
vasishth@uni-potsdam.de

My comments on the "noisy channel" model

- The "noisy channel" model as described by Traxler seems like a vague hand-waving story reminiscent of the 1980s.
- It reminds me of Relevance Theory—whatever result you get you can explain it in terms of the powerful language of Bayes Theorem.
- It backs away from a process model, allowing the researcher lots of degrees of freedom in explaining pretty much any outcome.
- A nice test of a theory is: what is a counter-example? For the "noisy-channel" model it is going to be hard to come up with one.
- So my take on the "noisy channel" model is that it tries to subsume all other theories by backing off to a level of generality that is not particularly useful.
- Process models are much more useful because they force the researcher to make a commitment.

Shravan Vasishth
vasishth@uni-potsdam.de

My comments on the "noisy channel" model

- Consider the complex issues raised in the 69 studies reviewed in Engelmann et al last week. Traxler's review references Gordon et al's studies on subject vs object relatives when discussing how SBI can be explained in terms of the "noisy channel" idea. I find this extreme reductionistm quite amazing, even for a review article.
- I guess I can explain the diverse results in the Engelmann review as follows:
  - In target match condition, we might be facilitatory, inhibitory or no interference depending on the extent to which the reader backs off to his priors.
  - In target mismatch condition, we might be facilitatory, inhibitory or no interference depending on the extent to which the reader backs off to his priors.

  At the level of generality at which the "noisy channel" model is stated, we can always instantiate in a way to fit the data without gaining much insight.

(Note that Levy did implement the model in his original paper; his approach seems to have fallen by the wayside in favor of hand-waving accounts).

Shravan Vasishth
vasishth@uni-potsdam.de

GE parsing and the "noisy channel" model

Traxler considers the "noisy channel" model compatible with GE parsing.

"Because comprehenders? interpretations in these studies were systematically biased toward plausible meanings, these findings offer a clear demonstration of the effects of prior knowledge and beliefs. That is, they indicate that anomalous interpretations are not adopted even when the syntax of a sentence provides clear indications that the anomalous meaning was intended. Thus, the results are straightforwardly compatible with the Bayesian/noisy-channel approach to sentence interpretation."

I don't know what to make of this, as pretty much everything is compatible with Bayes Theorem, once you set it up in terms of Bayes.

Shravan Vasishth vasishth@uni-potsdam.de

Underspecification

See Traxler Fig 1.
Logacev et al 2015a (Cog Sci), 2015b (QJEP) have tried to develop models of underspecification in a more formal way than the box and arrows model in Fig 1. I obviously feel that is the right way to investigate underspecification theories—by implementing them computationally.

Individual differences

Traxler makes the important observation that working memory resources may impact prediction.
Bruno Nicenboim's work explores this. He is also exploring the role of cognitive control in inhibitory processes (in highly predictive and low predictive situations).

Shravan Vasishth
vasishth@uni-potsdam.de

Concluding remarks

- There seems to be a general tendency to exaggerate the newness of ideas, rather than reconnecting them with older investigations.
- As a result, older work gets lost and connections are missed (e.g., the prediction industry in psycholinguistics has missed the details on the possible role of inhibition in predictive contexts).
- The data are much less clear than the review suggests. This is due to a general overconfidence about what we think the data tell us (a major culprit is the lack of statistical education, and the terrible distortions that psychology introduced into statistical inference).
- We should be moving towards implemented computational models in 2015; instead, the review is quite encouraging towards hand-waving models.

Shravan Vasishth
vasishth@uni-potsdam.de

Concluding remarks

The review does lay out one perspective on recent work on what one might call "underspecification" driven by prior beliefs. The connection made to surprisal type metrics seems rather vague in this review.
Spelling out these vague ideas would make for some great PhD dissertations.

Shravan Vasishth
vasishth@uni-potsdam.de