Fixation durations in first pass reading reflect uncertainty about word identity

Nathaniel J. Smith <njsmith@cogsci.ucsd.edu>
UC San Diego

Roger Levy <rlevy@ling.ucsd.edu>
UC San Diego

What kind of input does the language processor receive?

Noise is an inevitable component of sensory systems and biological computation generally. Yet many psycholinguistic theories assume that the identity of the words to be processed is known exactly. Is this true?

Option 1: Ambiguity about word identity is resolved early and quickly, before deeper linguistic processing occurs.
→ *Predicts* High-level linguistic properties of words - like surprisal - can affect processing directly.

Option 2: Ambiguity about word identity is preserved and propagated by the processing system.
→ *Predicts* Processing **cannot** be affected by word properties like surprisal *per se*, because the word itself is unknown.

But if surprisal doesn't affect processing, what does?
We define a new quantity, *average neighborhood surprisal*, and show that it explains fixation times better than raw surprisal. Readers look longer at words with high surprisal neighbors, and look shorter at words with low surprisal neighbors.

Previously...

Ehrlich and Rayner (1981): When reading a misspelled word, if the context makes the correct word predictable then readers are less likely to fixate the misspelling, fixate for marginally shorter times, and are less likely to report noticing it.

Connine, Blasko, and Hall (1991): Listeners are willing to revise their identification of perceptually ambiguous phonetic material in light of disambiguating material that follows within a short period.

Norris (2006): Argues that the reason high neighborhood density speeds lexical decision but slows naming is that lexical decision does not require unambiguous word identification, but naming does.

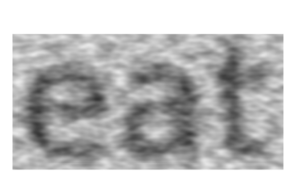
Levy, Bicknell, Slattery, and Rayner (2009): Difficulty with locally coherent sentences is reduced by eliminating *neighbors* that otherwise introduce parsing ambiguity.

Levy (tomorrow, poster 2.38): Readers boggle at garden paths that aren't there.

This poster: a general approach to accounting for uncertain input in models of word effects.

Model: intuition

They petted the...



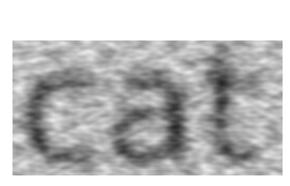
cat — Surprisal = 1; $P(\textbf{cat} \mid \text{input, context}) = 0.1$

eat — Surprisal = 10; $P(\textbf{eat} \mid \text{input, context}) = 0.9$

Prediction: *faster* than under raw surprisal

Average **N**eighborhood **S**urprisal $= \mathbf{E}(\text{Surprisal}) = 1 \times 0.1 + 10 \times 0.9 = 9.1$

They petted the...



cat — Surprisal = 1; $P(\textbf{cat} \mid \text{input, context}) = 0.95$

eat — Surprisal = 10; $P(\textbf{eat} \mid \text{input, context}) = 0.05$

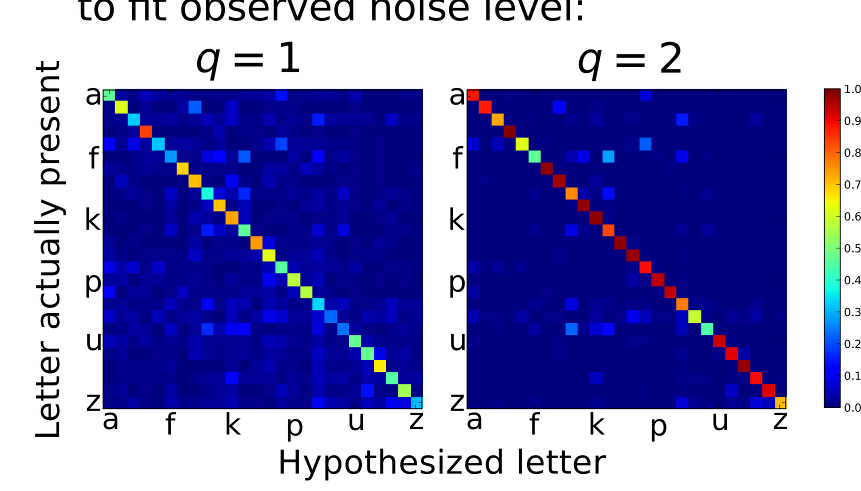
Prediction: *slower* than under raw surprisal

Average **N**eighborhood **S**urprisal $= \mathbf{E}(\text{Surprisal}) = 1 \times 0.95 + 10 \times 0.05 = 1.45$

Model: math

Bayesian model of word recognition under visual uncertainty

Model visual noise as uncertainty about *letter identity*
Quantify with a confusion matrix (from norming studies: Engel, Dougherty, and Jones, 1973; Geyer, 1977).
Free parameter q rescales confusion matrix to fit observed noise level:



For letters:

$$P(\text{input} \mid \text{letter}) = \frac{P(\text{letter} \mid \text{input}) P(\text{input})}{P(\text{letter})}$$

(From (rescaled) confusion matrix; Assumed uniform)

For words:

$$P(\text{word}_i \mid \text{input, context}) = \frac{P(\text{word}_i \mid \text{context}) P(\text{input} \mid \text{word}_i, \text{context})}{P(\text{input} \mid \text{context})} = \frac{P(\text{word}_i \mid \text{context}) \prod_j P(\text{input}_j \mid \text{word}_{ij})}{Z}$$

(Conditional independence; Top-down prior from context; Letter model from above)

Average neighborhood surprisal

We predict that reading times will be driven not by raw surprisal (RS), but by average neighborhood surprisal (ANS).

Raw surprisal $= RS_i = -\log P(\text{word}_i \mid \text{context})$ (Standard definition)

$$ANS = \mathbf{E}_i(RS_i) = \sum_i RS_i P(\text{word}_i \mid \text{input, context})$$

(Calculated by the Bayesian model)

Results

Method: Regressed fixations from the Dundee corpus against RS and ANS, with frequency and word length as control variables. RS and top-down prior estimated with Kneser-Ney smoothed BNC trigrams. Free parameter q fit simultaneously by maximum likelihood.

First fixations

$q = 1.3$
ANS: $t(182155) = 4.164, p \ll 0.001$
RS: $t(182155) = 0.489$, *n.s.*

Not much visual noise
Mean letter naming accuracy: 66%
→ ANS and RS highly correlated: $R^2 = 0.96$

But ANS still unambiguously beats RS ✔

Second fixations

$q = 2.9$
ANS: $t(42010) = 4.209, p \ll 0.001$
RS: $t(42010) = -1.847, p = 0.06$

Similar overall result, but note: less visual noise affecting second fixations, as we would expect (noise $= \frac{1}{q}$). ✔

RS marginally significant, but going the wrong way

Does context really matter?

$P(\text{seen} \mid \text{context}) \rightarrow P(\text{seen})$ Test a variant of ANS: for top-down prior, use simple word frequency instead of context-sensitive prior.

Context-sensitive ANS remains significant:
$t(182154) = 4.413, p \ll 0.001$
→ People really do marginalize over this complex prior. ✔

But, a mystery remains...

Can define "average neighborhood X" for any psycholinguistic property X, and make the same prediction.

We tried frequency. Regression preferred *raw* frequency.

Average neighborhood frequency with $q = 0$ does account for some additional variance, but our theory cannot explain such a small q.

? ?

Acknowledgements

We are grateful to Michael Tanenhaus for the initial suggestion of averaging surprisal over the visual neighborhood, and to Shane T. Mueller for maintaining the invaluable Letter Similarity Data Set Archive. This research was supported by NIH Training Grant T32-DC000041 to the Center for Research in Language at UC San Diego.

So...

We found broad-coverage evidence that the processor is able to maintain and efficiently marginalize over noise in its input, and dissociated frequency and surprisal effects. And we've gained some new questions, with perhaps the most urgent being:

Is this really visual noise, or something internal? (Next step: try a phonological noise model)
What's going on with frequency?