Final lecture

Shravan Vasishth

16 Dec, Tokyo

Introduction

The story so far:

- We began by reviewing some claims about retrieval-based processes, mainly basing the story on the Lewis and Vasishth 2005 model. This was the Engelmann et al paper. **This establishes the role of retrieval based processes in sentence comprehension. One open question is the relationship between retrieval and predictive parsing.**
- Then we read the Traxler review, which tried to bring together the recent work involving surprisal, noisy channel, and good enough processing. **I criticized the vagueness of the noisy channel model. Formalizing the idea as a Bayesian inferencing problem is a great research opportunity (Roger Levy has worked on this).**

Introduction

The story so far (continued):

- Then we looked at the evidence for surprisal from Levy and Keller 2013 and Jäger et al 2015 and Linzer and Jaeger 2015 (strong) and the evidence for entropy and related ideas (very weak to non-existent IMO). **Open research question: is the relevant metric Entropy or ERH? Is there any solid evidence for Entropy or ERH? How do these metrics compare with surprisal?**
- Next, we did a small excursus on the processing of ungrammatical embeddings: the English-German differences in RT could be explained in terms of surprisal-based theories (but Bader's work adds a nuance to that claim). We could dismiss this line of work in English entirely as being about totally weird sentences; but not so easy to do this for German. **The German and Dutch phenomena discussed need a lot more investigation, possibly from the silent prosody perspective.**

Introduction

The story so far (continued):

- Then, we tried to replicate Levy and Keller 2013's central claim about the interplay of surprisal and working memory load, and failed miserably. This may be just bad luck, or it may be an example of the general non-replicability of results in psycholinguistics. We can (and will) answer this question through repeated replication attempts. **Open research topic: can we replicate the key claims of surprisal, possibly with Japanese. Some surprises will be in store there.**

Note that non-replicability is not just a problem in psycholinguistics. It is a problem in science in general. See for example:
`http://socialbat.org/2015/08/12/goals-of-science-vs-goals-of-scientists-a-love-letter-for-`

Introduction

The story so far (continued):

- Also, we looked at Omaki et al's claim about how much detail is predicted about the upcoming verb. The data Omaki et al present seems to have an easy explanation in surprisal. **Nice thesis topic: Is hyper-active gap-filling the right way to think about it or is it really just surprisal?**

E.g., one could set up a situation in the Omaki design where surprisal (computed using sentence completions or corpus data) is constant, but transitives are predicted in both conditions that Omaki used.

Introduction

- Today we will look at another perspective on prediction, this one also from Colin Phillips' lab.
- It's called the bag of arguments view by the authors.

The basic idea is that the parser may initially use just the lexical association between arguments and verbs to predict, ignoring structural information (word order, or case marking)

suspect, police -> arrest

Introduction

The main paper we will discuss is

- Chow et al 2015 A "bag-of-arguments" mechanism for initial verb predictions

However, I will also make some digressions

- Momma et al UPenn slides (thanks to Colin Phillips)
- Husain et al 2014, Strong Expectations Cancel Locality Effects: Evidence from Hindi (downloadable from my home page)
- Chow et al (2014) Timing is everything: the temporal dynamics of word prediction (downloadable from Colin Phillips' home page)

Bag of arguments (Chow et al 2015)

Van Berkum et al 2005

The burglar had no trouble locating the secret family safe. Of course, it was situated behind a big_{NEU} / big_{COM} but unobtrusive painting_{NEU} / bookcase_{COM}.

Positivity when gender mismatches predicted noun (paintings).

Bag of arguments (Chow et al 2015)

Research question: Do all sources of information impact prediction immediately?

Chow et al's answer: No.

Main thesis

- "In the current study we examine how comprehenders use the meaning and structural position of the preceding words, in conjunction with their event-based knowledge, to compute lexical semantic predictions about an upcoming verb in a sentence."
- "We find evidence that the meaning of the arguments of the upcoming verb (rather than all preceding nouns) is rapidly used to predict that verb, but that the structural position of those arguments is used in bottom-up integration but is not as quickly incorporated into predictions."

Main thesis

- Prediction of verb depends on verb arguments
- But the structural position is not used in the early stages for prediction
- Does this go against Levy and Keller, who argue that the noun with dative case sharpens the prediction immediately?

Example

- Which cat did the landlord evict?

eviction event improbable because cat has role agent and is an improbably evictor

- Which landlord did the tenant evict?

eviction event improbable because tenant has role agent and is improbable evictor, and landlord is improbable evictee.

N400 and predictability

N400 Amplitude is inversely related to predictability

may reflect “facilitated long-term semantic memory access due to pre-activation of likely upcoming words or semantic features (Federmeier and Kutas, 1999)”.

Kukona et al 2011

"...listeners cannot immediately use information about the thematic role of a preceding argument to predict an upcoming argument."

Kukona et al's visual world study showed fixations to policeman at the verb in the passive:

1. Toby was arrested by the policeman.

but equal fixations to policeman and crook in (2)

2. Toby arrests the crook.

Conclusion: thematic role information is slow to be deployed to make predictions.

Chow and Phillips 2013

"Argument role reversals do not modulate N400 at verb"

the thief that the cop arrested

vs

the cop that the thief arrested

[Digression to Chow and Phillips paper]

More results consistent with Chow and Phillips null results

"Hoeks, Stowe, & Doedens, 2004; Kolk, Chwilla, Van Herten, & Oor, 2003; Van Herten, Chwilla, & Kolk, 2006; Van Herten, Kolk, & Chwilla, 2005)"

"These results constitute clear exceptions to the well-established generalisation that the N400 is modulated by a word's cloze probability."

From Chow et al 2014: "many recent studies across different languages (Dutch, Mandarin Chinese, Japanese and English) have reported that comprehenders' N400 response at the verb is unaffected by role-reversal of the pre-verbal arguments (Kolk, Chwilla, Van Herten & Oor, 2003; Hoeks, Stowe & Doedens, 2004; Van Herten, Kolk & Chwilla, 2005; Van Herten, Chwilla & Kolk, 2006; Ye & Zhou, 2008; Oishi & Sakamoto, 2010; Chow & Phillips, 2013)."

(Aside: Ye and Zhou 2008 do show an N400 in high cognitive controls)

Chow and Phillips 2013

They also argue that previous work (and their own work) has shown argument reversals cause a P600 effect, which

"...shows that comprehenders incrementally process information about the arguments' roles to interpret these sentences"

It's important to provide enough time to process the role information (see Chinese examples 1 in their paper)

Momma, Sakai and Phillips (in prep)

Momma et al replicated the Chow et al and found that

"merely increasing the SOA between an argument and a verb in simple two-word Subject-Verb and Object-Verb sentences revived the N400's sensitivity to argument role reversals."

Conclusion:

the structural roles of preverbal arguments can impact comprehenders' lexical semantic prediction about an upcoming verb only after a delay.

Garnsey et al 1989

They found N400 effects appearing due to role implausibility:

The businessman knew which customer/article the secretary called. . .

A larger N400 at *called* in article vs customer conditions suggests early use of animacy information.

Conclusion: some information is used early.

Summary of previous literature

"...these results suggest that comprehenders can use both the meaning and structural roles of preceding words to pre-activate likely upcoming verbs (or their lexical semantic features), but the predictive processes involved are initially sensitive to only a subset of this information"

Goal of present study, and the main claim

"we investigate exactly which information is recruited in the initial stage of verb prediction and why."

Claim:

- When nouns are processed, identify arguments and predict verb based on this. (Bag-of-arguments)
- Structural roles are not used for prediction, but some rudimentary syntactic chunking happens

Implication: verb predictions are informed by clausal structure of the preceding sentence.

Experimental manipulations

- Argument role reversal

The restaurant owner forgot which customer the waitress had served during dinner yesterday.

The restaurant owner forgot which waitress the customer had served during dinner yesterday.

- Argument substitution

The superintendent overheard which *tenant* the landlord had evicted at the end of May.

The superintendent overheard which *realtor* the landlord had evicted at the end of May.

Experimental manipulations

My Comments

- Notice that in the role reversal conditions there is a potential for spillover effects confounding the results. I would not consider the two factors a fair comparison.
- Spillover confounds have been a thorn in our side for many years now.
- A better design is needed.

Expt 1

Tests sensitivity to lexical meaning and structural roles when making predictions of a verb

Expt 1 cloze task

"In the high cloze conditions the target verbs had an average cloze probability of 25% (range = 13–53%) in the argument role reversal sentences and 28% (range = 13–77%) in the argument substitution sentences."

Research opportunity: I find the cloze scores too low to count as predictable. What happens when we have high predictability (in the range of 70-80%)?

Expt 1 predictions

Partly quoting them directly:

- Based on previous findings, we expect the N400 response at the target verb to be insensitive to cloze probability differences that resulted from argument role reversals.

My Comment: to test this, we could use cloze probability for individual verbs as a predictor in the model.

- However, if comprehenders can quickly use the lexical meaning of the arguments for verb predictions, then we should observe an N400 effect at the target verb in the argument substitution conditions, despite the fact that the size of the target cloze probability difference in the argument substitution conditions was highly comparable to that in the argument role reversal conditions
- P600 in implausible conditions.

Latent Semantic Analysis

LSA was used to compute similarity values between pairs of nouns.

"the verb was mildly associated with the fronted object in both high and low cloze conditions, with mean SSVs of .22 (SD = .2) and .12 (SD = .12) in the high and low cloze conditions respectively."

Experiment task

Each sentence was preceded by a fixation cross that appeared for 500 ms. Each word appeared on the screen for 300 ms, followed by 230 ms of blank screen (i.e. 530 ms SOA). The last word of each sentence was marked with a period, followed 1000 ms later by the question "Is this sentence plausible?"

Note that this task could encourage superficial processing; Hoeks et al and Kolk et al also had judgement tasks. Ye and Zhou 2008 had comprehension questions and found an N400 in high control subjects.

Filler characteristics

Quoting directly

- Sixty pairs of argument role reversal sentences and 60 pairs of argument substitution sentences were distributed in two presentation lists, such that only one member of each pair appeared in each list (30 per condition).
- In addition, each list also contained 140 filler sentences. Sixty of these filler sentences were included to test the effects of cloze probability in strongly constraining sentences (cloze probability of high cloze words $>75\%$) and served as a control comparison for the current experiment.
- These sentences were adapted from a previous experiment examining the effects of cloze probability: each sentence context had a high cloze target and a low cloze target as a continuation, and the high cloze target in one context appeared as the low cloze target in another context.

Filler characteristics

Quoting directly

- These sentences were distributed in two presentation lists such that each participant read 30 sentences with a high cloze target and 30 sentences with a low cloze target.
- An additional 80 unrelated filler sentences with similar length and structural complexity were included in both presentation lists.
- The overall high-to-low cloze ratio in each presentation list was 1:1.

Results Expt 1

See Figs 2 and 3

Only argument substitution affected N400 (interaction between context and cloze).

So, only the lexical meanings of the arguments played a role in prediction, not their structural roles.

My comments

- This seems related to the good-enough processing idea.
- The result may be limited to English; in languages with case-marking, there are many more cues as regards structural role
- An obvious implication is that in languages like Japanese and German and Hindi, we may see early use of structural information. (Colin disagrees)

Expt 2

Does the prediction process selectively use the words in the current clause to predict the upcoming verb?

- Is the prediction made using all nouns:
- exterminator - neighbor - landlord -> predicted verb: evicted;
- Or is the prediction made using only the current clause nouns:

exterminator - neighbor - landlord -> predicted verb: evicted

My comment: This logic seems to ignore the possibility that the first noun could be retained much more poorly in memory. Also recall Tabor et al study. They discuss this later in the paper.

Expt 2 Discussion

"The current results are not compatible with our formulation of the Bag-of-words hypothesis, according to which all preceding words (arguments or not) impact comprehenders' predictions in a uniform, bag-of-words fashion. Under this account, argument substitution should have no effects on the N400 at the verb as long as the verb is preceded by the same set of words in the sentence, as in the same-words conditions in the current experiment."

Expt 2 Discussion

Bag-of-words + decay has support.

General Discussion and my comments

- Arguments are quickly isolated and verbs that are thematically related to the arguments are predicted
- Structural roles are not used in the earlier stages

My comments

- In Hindi, we have evidence that case-marking information (structural role information) is used very quickly to generate predictions (Husain et al 2014, PLoS ONE, Expt 1)
- If this is correct, it's not clear to me why case-marking information would not be used in other situations.
- This seems like an important open question.