

How do words inform? Explaining the role of information theory in language comprehension

Final Summary Report

Words of a sentence or text are cognitively more difficult to process if they convey more information. But what does it mean to “convey information”? The discipline of Information Theory provides us with formal means to precisely define the amount of information conveyed by a word. And in the field of Computational Linguistics, models have been developed that can actually estimate these information values for each word of a text. When these values are compared to the time it takes to read the text, it turns out that information and reading time (as a proxy for cognitive processing load) are clearly related. The question remains, however, why there would be such a relation between information-theoretic quantities and cognitive processing. An answer to this question will teach us a lot about the human cognitive system for language comprehension, and may even bear upon the mystery of why languages are the way they are.

The main objective of the project was to uncover the relation between information and comprehension by looking at cognitive processing measures that go beyond the reading-time data that has been studied so far. Brain activity, too, is indicative of cognitive processing load, but unlike reading time, brain activity is *multidimensional*, which is to say that brain activities resulting from a piece of language vary not only in their *amount*, but also in the brain *areas* involved and the *moments* at which they become active.

Method

The basic methodology is to perform large-scale analyses to find correlations between (1) model-derived quantities of the amount of information conveyed by words in sentences, and (2) measures of human language processing load during comprehension of these sentences. The models learn to recognize statistical patterns from very large text corpora and are then able to estimate the occurrence probabilities of words given their sentence context. Information values are derived from these probabilities. Different models implement different assumptions about the types of representations and processes involved in language comprehension.

The models’ predictions are compared to human processing data. Three types of data were collected from people engaged in language comprehension: Eye movements were registered during reading, and brain activity was measured by EEG during reading as well as by fMRI during listening to spoken texts.¹ The language stimuli used in these studies are not hand-crafted for the sake of the experiment but sampled from sources that were written for other purposes. That is, the stimuli were naturalistic texts. More specifically, they were sentences or short excerpts from novels. This means that, if the models reliably predict processing measures in these materials, this result is very likely to generalize to different materials. In contrast, applying a particular experimental manipulation to unnatural stimuli (the common approach in psycholinguistics) may yield effects that are limited to this particular manipulation or these particular items. The use of naturalistic stimuli for this project is facilitated by the application of state-of-the-art models of language, which are able to process (and generate information values for) any sentence. Crucially, different models (embodying different assumptions about the language system) will estimate different information values, so comparisons between the different models’ predictions can uncover the cognitively most plausible representations or mechanisms.

Main findings

- The language models are able to account for brain activity during comprehension of naturalistic text. This has previously only been demonstrated on behavioural (i.e., word-reading time) data.
- Information measures based on models that include linguistically informed structures (i.e., probabilistic phrase-structure grammars) do not predict the N400 EEG response any better than simpler models do.
- Two information-theoretic notions that (informally put) measure the *unexpectedness* of a word’s occurrence and the *uncertainty* about which word(s) may come next, both account for brain activity but

¹ The collection of the two sets of neuroimaging data was not part of the current project but funded through other means.

show effects in different brain regions. Interpretations of these regions suggest that being in a state of (relatively) high certainty leads to the active prediction of the form up the upcoming word.

- Cognitively, the semantic relatedness between words of a sentence or text cannot be reinterpreted in terms of word predictability. Model-derived measures of relatedness and predictability lead to different brain responses, as visible in the fMRI data. However, the EEG data suggest that both relatedness and predictability simultaneously affect the extent to which words are anticipated.
- The eye-tracking data revealed that fluent readers of a second language are affected by word information values in the same way as native speakers.

Main conclusions

The findings support a recent theory of cognitive and neural processing known as the “Hierarchical prediction framework”. According to this idea, the brain is continuously generating “top-down” predictions about the upcoming stimuli. That is, knowledge and context information are used to pre-activate representations of the possible inputs. The activations are graded, in the sense that inputs are pre-activated relative to their probability of occurrence. To the extent that the actual input *mismatches* the pre-activation, an “error signal” is propagated “bottom-up”, which in turn can alter subsequent predictions.

The finding that modelling a sentence’s syntactic structure does not lead to a better fit to brain activity than using simpler representations suggests that the brain/mind does not make use such structures, or at least that not all available linguistic knowledge is applied to generate anticipations about upcoming words.

Potential impact

The project sits at the forefront of two current trends in the cognitive sciences: One towards the use of more naturalistic stimuli (both in behavioural and neuroimaging studies), and the other towards a tight integration of experimental and computational research methods. These trends are likely to have major impact on the scientific field in the coming years. The successful application of model-derived quantities to neuroimaging data from natural sentence/text comprehension studies was, to the best of our knowledge, the first in the published literature. This method opens up new possibilities of running neuroimaging experiments and using the collected data. More specifically, this approach can complement traditional experimental designs and lead to results that generalize better to other stimuli and are less prone to being affected by participants’ strategic adaptation to the manipulations or structures of the experimental items. Additionally, the novel finding that the combination of naturalistic stimuli and model-based word-level predictions also works when participants read in their second language means that the same methods can be applied in research on bilingualism and non-native language processing. It may therefore also have a major impact on those subfields of psycholinguistics as well.

Contact details

Stefan Frank

Centre for Language Studies

Radboud University Nijmegen

s.frank@let.ru.nl

www.stefanfrank.info

Tel.: +31 24 3615491