Computational Cognitive Science

Lecture 16: Sentence Processing 1

Frank Keller

12 November 2026

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Overview

Sentence Processing

Expectation and Processing Difficulty

Surprisal

Empirical Evaluation

Assessment

Reading: Hale (2016).
Optional: Slaats and Martin (2025).

Sentence Processing

What Is Sentence Processing?

Language comprehension involves more than recognizing individual words.

words $\Longrightarrow$ structured interpretation

Sentence processing (or parsing) is the process of determining how the words in a sentence relate to one another.

For example:

(1) The kitten bit the dog.

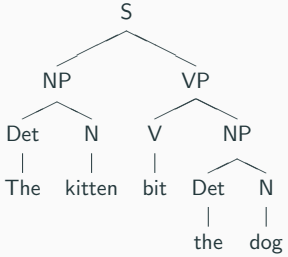
Sentence (1) is not simply a list of words: we understand *the kitten* as the agent and *the dog* as the object of *bit*.

Grammatical Structure

A simple way to represent sentence structure is with a context-free grammar:

S	$\rightarrow$	NP VP	NP	$\rightarrow$	Det N
VP	$\rightarrow$	V NP	Det	$\rightarrow$	the
N	$\rightarrow$	kitten, dog	V	$\rightarrow$	bit

These rules license a hierarchical representation:



Parsing means constructing structure from an incoming word sequence.

Syntactic Ambiguity

A sentence can be compatible with more than one syntactic structure.

Examples:

(2) a. She saw the man with the telescope.
b. I put the book on the table in the kitchen.
c. Kids make nutritious snacks.

This is global ambiguity: even after the entire sentence has been heard or read, more than one analysis is possible.

Human sentence processing must therefore do more than apply grammar rules: it must decide which analyses are more plausible.

Incremental Processing

The human sentence processor is incremental: interpretation is updated word by word as the input arrives.

Consider:

(3) The athlete realized his potential ...

a. ... at the competition.

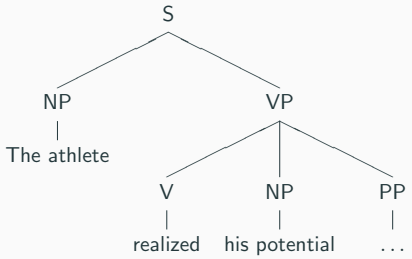
b. ... could make him a world-class sprinter.

At the point marked by ..., both continuations are possible.

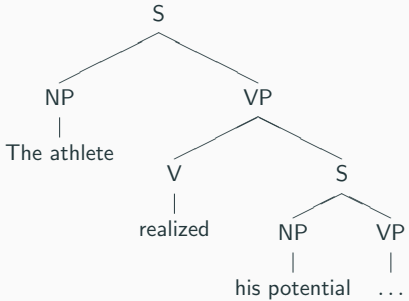
This is local ambiguity: a sentence prefix is compatible with multiple analyses, even though later input may resolve the ambiguity.

Two Analyses of the Same Prefix

NP analysis



Clause analysis



... at the competition.

... could make him a world-class sprinter.

Before the disambiguating word arrives, both structures are possible.

Garden Paths

Sometimes the processor strongly favors an interpretation that later turns out to be wrong:

(4) The horse raced past the barn fell.

At *raced*, the preferred interpretation treats *raced* as the main verb. At *fell*, that interpretation becomes impossible.

Garden paths are associated with:

- longer reading times;
- regressions in eye movements;
- reduced comprehension accuracy.

Why are some locally ambiguous sentences much harder than others?

Not All Ambiguities Cause Garden Paths

Compare:

(5) a. The horse raced past the barn fell.
b. The bird found in the room died.

Both contain the same kind of temporary syntactic ambiguity. But the first produces a much stronger garden-path effect.

Similarly:

(6) a. The complex houses married and single students.
b. The warehouse fires a dozen employees each year.

Syntactic ambiguity alone is not enough to predict processing difficulty.

Expectation and Processing Difficulty

Expectations in Sentence Processing

There is direct evidence that readers form expectations about upcoming words.

Staub and Clifton (2006) compared:

(7) a. Peter read either a book or an essay in the school magazine.
b. Peter read a book or an essay in the school magazine.

The region following *or* is read faster when *either* has already appeared.

Hypothesis: this is because *either* makes *or* more expected.

Garden Paths as Violated Expectations

The same intuition applies to garden paths.

(8) a. The horse raced past the barn fell.

b. The bird found in the room died.

In (8-a), *fell* is highly unexpected after the preferred interpretation of the prefix. In (8-b), the reduced-relative interpretation is more plausible, so *died* is less unexpected.

Likewise:

(9) The complex houses married and single students.

Here, *married* violates a strong expectation that *houses* is the main verb.

Beyond Ambiguity

Processing difficulty also occurs in sentences that are not locally ambiguous.

Compare subject and object relative clauses:

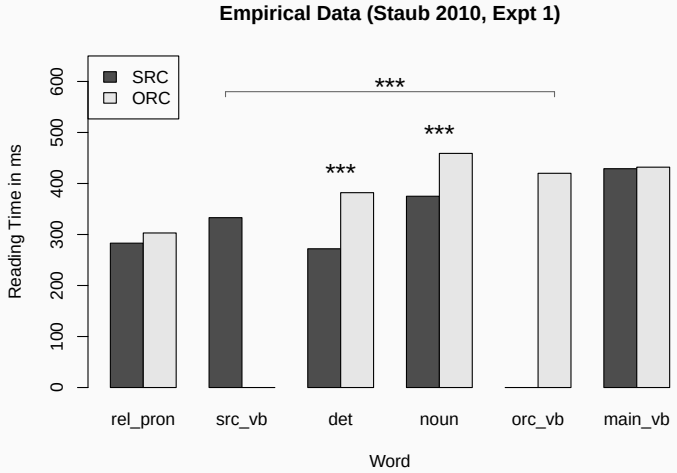
(10) a. The reporter that attacked the senator admitted the error.

b. The reporter that the senator attacked admitted the error.

The object relative clause in (10-b) is harder to process than the subject relative clause in (10-a).

We need a measure of processing difficulty that applies beyond garden paths and ambiguity.

Human Reading Times



Object relative clauses produce longer reading times at several regions (Staub 2010).

Can a general expectation-based measure predict such differences?

Time for a short quiz on WooClap!

Go to wooclap.com

Enter the event code in the top banner

Event code

GYMBIPL



Surprisal

Surprisal: Recall from Lecture 13

We already encountered surprisal when discussing active learning:

$$I(x) = -\log P(x)$$

An event carries more information when it was less probable:

Probability	Surprisal (bits)
1	0
1/2	1
1/4	2
1/8	3

When we model sentence processing, we can apply the same idea to the next word.

Word Surprisal

For a word w_k in context $w_1, \ldots, w_{k-1}$:

$$\boxed{S_k = -\log P(w_k | w_1, \ldots, w_{k-1})}$$

- High-probability word $\Rightarrow$ low surprisal.
- Low-probability word $\Rightarrow$ high surprisal.

The surprisal hypothesis predicts:

$$\boxed{\text{higher surprisal} \quad \Longrightarrow \quad \text{greater processing difficulty}}$$

Processing difficulty can be measured using reading times, eye movements, or other behavioral measures.

A Simple Example

Suppose that after the context

(11) She spread the warm bread with . . .

a model predicts:

Next word	Probability	Surprisal
butter	.50	1.0 bits
jam	.25	2.0 bits
cheese	.125	3.0 bits
socks	.001	10.0 bits

Surprisal makes quantitative predictions of processing difficulty.

Surprisal and Probability Mass

A syntactic interpretation of surprisal starts with all analyses compatible with the current prefix.

When a new word arrives:

- some analyses remain possible;
- others become incompatible and are eliminated.

If the new word eliminates a large amount of probability mass, it has high surprisal.

$$\text{many expected analyses ruled out} \quad \Rightarrow \quad \text{high processing cost}$$

This gives a probabilistic account of garden-path difficulty without assuming an all-or-none commitment to a single parse.

Prefix Probabilities

Let $P(w_1, \ldots, w_k)$ be the probability of the sentence prefix up to word k. Then:

$$P(w_k | w_1, \ldots, w_{k-1}) = \frac{P(w_1, \ldots, w_k)}{P(w_1, \ldots, w_{k-1})}$$

Therefore:

$$\boxed{S_k = -\log \frac{P(w_1, \ldots, w_k)}{P(w_1, \ldots, w_{k-1})}}$$

Surprisal measures how much probability mass is "lost" when the prefix is extended by the next word.

From Syntactic Surprisal to Word Surprisal

In a probabilistic parser, the probability of a prefix is obtained by summing over all compatible analyses T:

$$P(w_1, \ldots, w_k) = \sum_T P(T, w_1, \ldots, w_k)$$

Substituting this into the previous equation gives word surprisal.

Importantly, this is independent of the analysis T (Levy 2008), hence:

$$S_k = -\log P(w_k | w_1, \ldots, w_{k-1})$$

The final definition does not require a particular syntactic representation.

This makes surprisal a very general linking hypothesis between probabilistic expectations and human processing difficulty.

How Can We Estimate Surprisal?

To compute $P(w_k|w_1, \ldots, w_{k-1})$ we need a probabilistic language model.
Possible models include:

- probabilistic incremental parsers;
- n-gram models;
- recurrent neural networks;
- transformer language models.

Different models may assign different probabilities, but the surprisal-to-processing prediction is the same.

This separates the linking hypothesis from the model that supplies the expectations.

Surprisal as a Cognitive Model

Surprisal makes two distinct claims:

1. The comprehender maintains probabilistic expectations about upcoming linguistic input.
2. Processing cost increases with the unexpectedness of the input.

A language model can therefore be evaluated as a cognitive model by asking:

Do its word probabilities predict human processing difficulty?

This is different from evaluating the model only by how accurately it predicts text (for example in NLP).

Empirical Evaluation

Testing the Surprisal Hypothesis

A typical evaluation proceeds as follows:

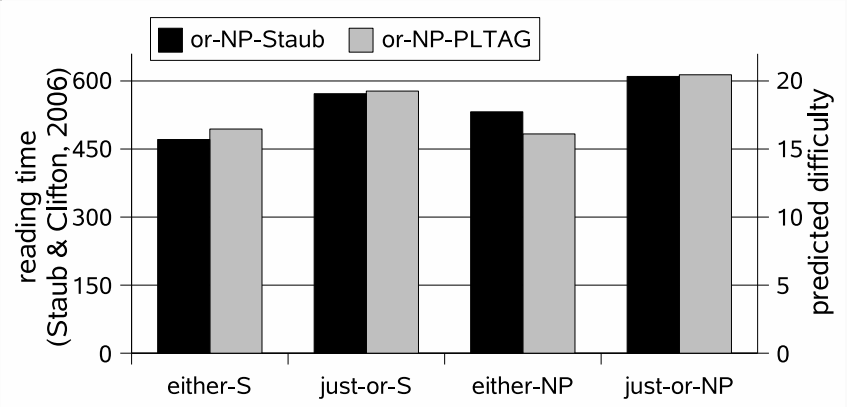
1. obtain sentences from a psycholinguistic experiment;
2. estimate word probabilities with a language model;
3. convert probabilities to word surprisals;
4. compare surprisal with human reading times or eye movements.

Prediction:

$$\text{higher surprisal} \quad \Rightarrow \quad \text{longer reading time.}$$

The comparison can be made for experimentally manipulated conditions or over large naturalistic reading-time corpora.

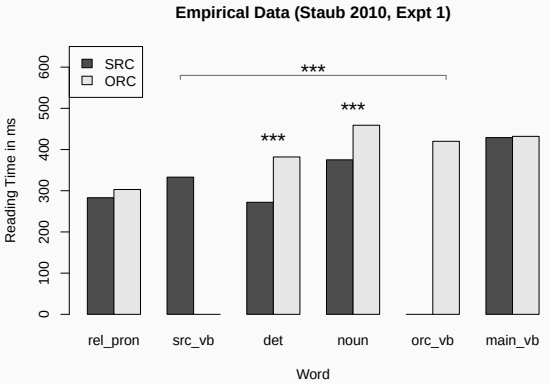
Result: either . . . or

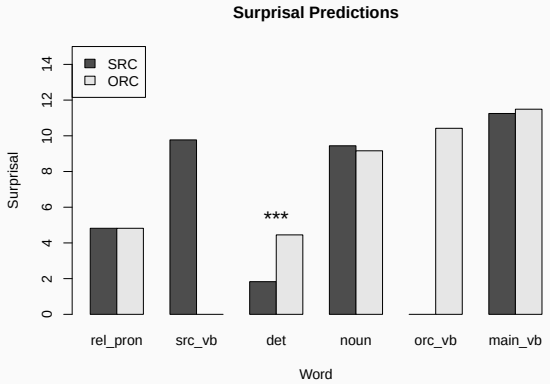


The expectation created by *either* reduces surprisal when *or* arrives, matching the reading-time facilitation.

More predictable continuations are processed more easily.

Result: Relative Clauses





Surprisal captures part of the subject/object relative-clause difference, but not all of it.

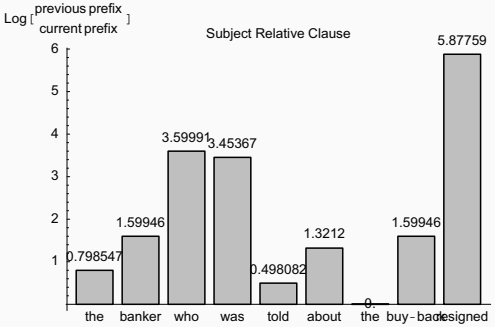
In particular, additional memory costs needed to explain difficulty associated with long-distance dependencies (Demberg et al. 2013).

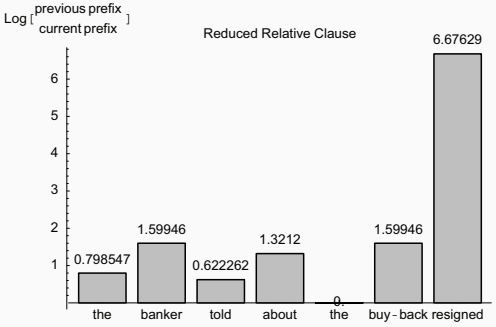
Result: Garden Paths

Compare a reduced and an unreduced relative clause:

(12) a. The banker who was told about the buy-back resigned.

b. The banker told about the buy-back resigned.





Disambiguating region in the garden-path sentence has higher surprisal (Hale 2001).

Assessment

Strengths and Limitations

Strengths

- one quantitative measure applies to many phenomena;
- naturally incremental;
- makes graded predictions;
- works with many probabilistic language models;
- connects directly to measurable reading times.

Limitations

- expectation is not the only source of processing difficulty;
- memory and retrieval costs matter;
- predictions depend on the probability model used;
- good text prediction does not guarantee cognitive plausibility.

Summary

- Human sentence processing is incremental: interpretations updated word by word.
- Garden paths show that processing depends on expectations, not just on whether a sentence is syntactically ambiguous.
- Surprisal is $S_k = -\log P(w_k | w_1, \ldots, w_{k-1})$. Higher surprisal predicts greater processing difficulty.
- Surprisal is representationally agnostic: probabilities can come from a probabilistic parser or a neural language model.
- Surprisal explains important reading-time effects, but memory and other processing mechanisms are also required.
- Next time: Modern extensions of surprisal, including using large language models.

References

Demberg, Vera, Frank Keller, and Alexander Koller. 2013. Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics* 39(4):1025–1066.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Pittsburgh, PA, volume 2, pages 159–166.

Hale, John. 2016. Information-theoretical complexity metrics. *Language and Linguistics Compass* 10(9):397–412.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3):1126–1177.

Slaats, Sophie and Andrea E. Martin. 2025. What's surprising about surprisal. *Computational Brain & Behavior* 8(2):233–248.

Staub, Adrian. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition* 116:71–86.

Staub, Adrian and Charles Clifton. 2006. Syntactic prediction in language comprehension: Evidence from either . . . or. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:425–436.