Difficulty of Sentence Comprehension

How information theory can help us understand which sentences are difficult to comprehend.

Topic of the Talk & Preliminary Remarks

- We will be discussing an information theoretic approach to predicting the difficulty of sentence comprehension.
- We will clarify a proposal by Roger Levy that we can predict the difficulty of a sentence by looking at the surprisal of a word in a sentence.
- No background knowledge in lunguistics is required. (I am also no linguist).
- Please feel free to interrupt me at any time.

Information Theory & Language

- Already Shannon and Weaver did research on langauge in information theoretic terms.
- Shannon and Weaver wanted to know what the redundancy of the English Language was.
- Redundancy is the fraction of the message which is determined not by free choice but by statistical rules.
- Shannon and Weaver wanted to know how many redundancy you need for satisfactory crossword puzzles.

Difficulty of Sentence Comprehension

- This talk will focus on the difficulty of comprehending a sentence.
- We are looking for an answer to the question: “What makes a sentence difficult to read?”
- We measure difficulty by reading time.
- Why do we want to know which sentences are difficult?
- So what properties of sentences or words make sentences difficult? Some intuitive answers come to mind very quickly.

Which sentences are more difficult to comprehend?

- **Is this determined by length?**
  1) The player that the coach met at 8 o'clock bought the house.
  2) The player that the coach met near the gym by the river at 8 o'clock bought the house.
- **Is this determined by grammatical structure?**
  1) The reporter who attacked the senator admitted the error.
  2) The reporter who the senator attacked admitted the error.
- **Is this determined by presence of ambiguity?**
  1) The daughter of the colonel who shot herself on the balcony had been very depressed.
  2) The son of the colonel who shot himself on the balcony had been very depressed.
- **Is this determined by the frequency of word occurrences?**
  1) We bought a coffee at the coffee shop on the corner.
  2) We bought a soya bean at the flag on the Kullback–Leibler-distance.

Sentence Comprehension

- Evidence suggest sentence comprehension is incremental, i.e. on line.
- Any formalization of sentence comprehension should accommodate incremental processing
- Natural to describe on line comprehension of sentence by way of **probability distributions of possible continuations of a sentence** based on the words we already read in the sentence.
- Before this is possible we need another definition.

Definition: Complete Structures

A language contains an infinite set of complete structures $\mathcal{T}$ such that a fully disambiguated utterance corresponds to exactly one structure.

Example: The girl saw the boy with a telescope

Now this sentence is ambiguous. It can be that the girl saw some boy by looking through a telescope, or that the girl saw a boy carrying a telescope. This means that the sentence corresponds to 2 complete structures. Most sentences correspond to just 1 structure.

Conclusion: complete structures and sentences are different things!

Definition: Comprehending a Sentence

An agent comprehends a partial input sequence $w_1,...,w_i$ (the first i words of a sentence) by constructing a probability distribution P_{Ti} over the possible structures T based on $w_1,...,w_i$, where T in $\mathcal{T}$. The probability distribution P_T is updated after each word read.

The probability distributions forces an allocation or resources: i.e. it determines how much thinking power should pursuit which structure T.

Intuitively: while reading a sentence, after each word we read, we update a probability distribution over possible complete structures that will correspond with the sentence as a whole.

Definition Comprehension Difficulty

The difficulty of the incremental comprehension of a sentence is the amount of relative entropy between P_{Tk} and P_{Tk+1} defined as:

$$D(P_{T_{k+1}} || P_{T_k}) = \sum_{T \in \mathcal{T}_{k+1}} P_{T_{k+1}}(T) \log \frac{P_{T_{k+1}}(T)}{P_{T_k}(T)}$$

Note that this is just the normal definition of relative entropy (or Kullback-Leibler-divergence or Kullback-Leibler-distance) also found in *Cover & Thomas (pp. 19-20)*. Remember that the relative entropy is always non-negative and it is zero iff two probability distributions are equal (and becomes greater if the difference between the two probability distributions is greater).

Where is surprisal?

- It turns out that our definition of sentence comprehension difficulty in terms of relative entropy is equal to surprisal. That is:

the predicted difficulty of the *i*th word, w_i, is precisely equal to the surprisal of w_i given $w_1,...,w_{i-1}$.

Definition Comprehension Difficulty

The difficulty of the incremental comprehension of a sentence is the surprisal value of the *i*th word given words $w_1,\ldots,w_{i-1}$, defined as:

$$-\log P_{T_{i-1}}(w_i|w_1,\ldots,w_{i-1})$$

Proof of Equivalence Suprisal and Relative Entropy

This last claim is not trivial. So we will prove that the surprisal of the *i*th word is equal to the relative entropy between the probability distributions of the *i*th and *i-1*th word. This means we will prove the following statement:

$$\sum_{T \in \mathcal{T}_{k+1}} P_{T_{k+1}}(T) \log \frac{P_{T_{k+1}}(T)}{P_{T_k}(T)} = -\log P_{T_k}(w_{k+1} | w_1, \ldots, w_k)$$

Implications of Equivalence Proved

- We can now explain comprehension difficulty in terms of surprisal or relative entropy, this is useful because:
  - ➔ we can use the framework of relative entropy for resource allocation;
  - ➔ We can use the surprisal framework to describe why an agent takes more time to read a certain word or sentence
- Surprisal is easier to calculate than relative entropy, hence computational benefits.

Experimental Results Sentence Comprehension 1

- Consider the following three sentences:

  1) The player that the coach met at 8 o'clock$_1$ bought the house.

  2) The player that the coach met by the river$_1$ at 8 o'clock$_2$ bought the house.

  3) The player that the coach met near the gym$_1$ by the river$_2$ at 8 o'clock$_3$ bought the house

Experimental Results Sentence Comprehension 1

- A theory concentrating on grammatical structure would predict that three sentences are ordered by their difficulty, s.t.:
  - first one easiest;
  - last one hardest.
- What would surprisal predict?

Table 3
Surprisal and average reading times at matrix verb for (6)

	Number of PPs intervening between embedded and matrix verb		
	1 PP	2 PPs	3 PPs
DLT prediction	Easier	Harder	Hardest
Surprisal	13.87	13.54	13.40
Mean reading time (ms)	510 ± 34	410 ± 21	394 ± 16

Experimental Results Sentence Comprehension 2

- Consider the following three sentences:

  1) The **daughter** of the colonel who shot **herself** on the balcony had been very depressed.

  2) The daughter of the **colonel** who shot **himself** on the balcony had been very depressed.

  3) The **son** of the **colonel** who shot **himself** on the balcony had been very depressed.

Experimental Results Sentence Comprehension 2

- In most processing theories local structural ambiguity leads to higher comprehension difficulty.
- Structural ambiguity plays only a role in the determination of processing difficulty for the suprisal theory, if the ambiguity effects the conditional word probabilities.
- So what are the reading times for the three sentences?

Conclusions from Experiments

- There is a lot more evidence in the paper by Roger Levy. There are some especially nice examples and results in German.
- Surprisal seems best in predicting reading times, and thus the best framework for describing sentence comprehension difficulty.

Take Home Message

- If we want to predict the reading difficulty of a sentence, i.e. the reading time, we should just look at the cumulative surprisal of a sentence
- The surprisal theory is language independent. It is not important which language we are studying, only the underlying probabilities of a language are important.

References:

- Levy, Roger - “Expectation-Based Syntactic Comprehension” - Cognition, 2008, Vol.106(3), pp.1126-1177.
- Shannon, Claude - “The Mathematical Theory of Communication” - The Mathematical Theory of Communication, 1949, University of Illinois Press, pp. 29-125.
- Weaver, Warren - “Some Recent Contributions to the Mathematical Theory of Communication” - The Mathematical Theory of Communication, 1949, University of Illinois Press, pp. 1-29.

Thank you!