Sentence Length

Sentence length: Definition

Sentence length is measured in number of tokens: words, punctuation marks, numbers. Sentence length is often linked to sentence complexity and text readability. While complexity, as measured for instance by the number of different clauses in a sentence, is likely to result in longer sentences, the contrary is not necessarily true. Not all longer sentences are complex or unreadable. There is nothing too complex about a sentence made up of a long string of nouns with a verb. One thing, however, is for sure: sentence length has been steadily declining over the last century to the current average sentence length of less than 20 words. As John O'Hayre writes in his *Goobledygook Has Gotta Go* (1966):

> It's not unusual to find sentences of 75 or 100 words in Winston Churchill's writings, yet he is considered one of the great writers of the last half century. It isn't fair, then, to arbitrarily impose a rigid word-count on any writer. Neither is it fair for the writer to ignore the great gobs of research which show that the average reader today, whether a high school or a college graduate, overwhelmingly prefers to read sentences that average out at around 20 words.

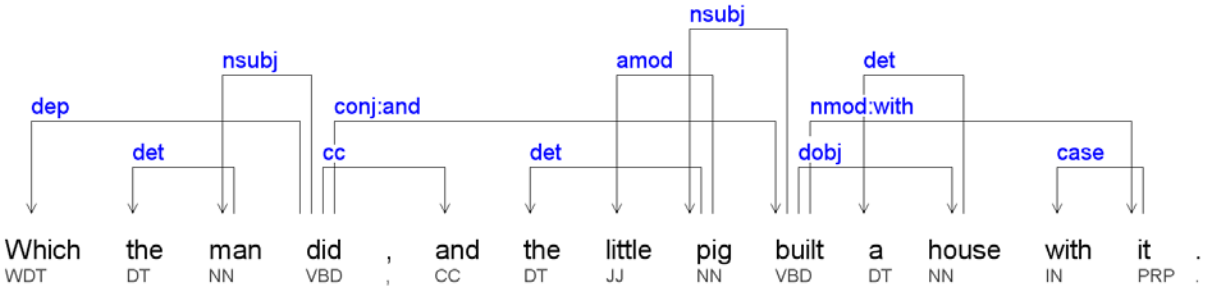
Other useful tools

Sentence complexity

You can compute measures of sentence complexity using the java script Sentence_Complexity.jar.

Visualizing the parse tree

Do try visualizing the parse tree using the java script dependenSee.jar.



Text readability

The text readability script provides several measures of readability, in terms of how many years of formal education it takes to understand a text, starting from grade 1 (of the American education system).

Unusual words

Unusual words (and not just because a word is misspelled) do not contribute to sentence complexity at the syntactical level but they do contribute to text readability at the semantic level. Run the NLTK algorithm to get a list of unusual words in your documents.

References

O'Hayre, John. 1966. *Goobledygook Has Gotta Go*. Washington, D.C.: U.S. Government Printing Office No. 0-206-141.