# The Magical Number Seven in Language

**Emmanuel M. Pothos** (pothos@psy.ox.ac.uk)

Department of Experimental Psychology, University of Oxford,
Oxford OX1 3UD, UK

## Introduction

In Miller's (1956) classic work, participants were asked to rehearse sets of items. It was shown that, on average, people were good at retaining up to about seven items, and also that the limiting factor of this "rehearsal span" was the number of items actually present, rather than the information value of the items. An adequate explanation for this limitation (dubbed the magical number seven), one of the most robust in the experimental psychology literature, is still elusive.

One of the puzzling aspects of the problem is the size of the STM span. Why seven and not thirteen, for example? So-called idiots savants display a seemingly unlimited STM capacity–reportedly they can memorize random numbers of 10 digits as easily as of 100 digits (e.g., Sacks, 1985). Thus, there does not appear to be a biological factor constraining the STM to this particular size. So the apparent conclusion would seem to be that STM has evolved to be limited in this way possibly to optimize some aspects of the learning process (cf. Elman, 1993).

In this work, this issue is investigated in the reverse way: We examine one of the major products of human thought–language–and look at whether the patterns of regularities in linguistic text reflect in any way this STM span.

## Corpus

We used the combined Sherlock Holmes novels from the European Corpus Initiative, Multilingual Corpus 1 (ECI/MC1) database, which resulted in a corpus of approximately 750,000 tokens, and about 23,000 types. All tagging, capitalization, as well as punctuation, was removed, so that the material was effectively considered a long string of words.

## Analysis

Let "range" be the number of words in between two given words, $x$, and $y$. For instance, a range of 1 will indicate that words $x$ and $y$ are separated by only 1 other word. We are asking whether our expectation of obtaining word $y$ at a particular location is affected by the knowledge that we have word $x$ in an earlier location. A measure of this is the mutual information between $P(x)$ and $P(y)$, the probabilities of obtaining word $x$ and word $y$ respectively. Mutual information indicates how much the uncertainty involved in expecting $y$ is reduced by knowledge that we have $x$, and is given by $\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$. For different ranges, $P(x,y)$ is the probability of having both words $x$ and $y$, separated by a number of words equal to the range. The average mutual information of all unique word pairs was computed for ranges 0 to 16, and these values are plotted in Figure 1.

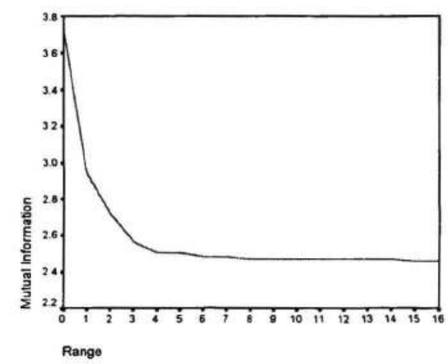


Figure 1.

As can be seen, the mutual information between different words levels off after about 7 words (this baseline is to be interpreted as the "zero point" mutual information). In other words, there are detectable statistical dependencies only between pairs of words separated by 7 or less other words.

Future work will address whether this kind of statistical regularity is evident also with other units of linguistic structure (for phonemes at least, this seems to be the case) and also the relation of the STM size with general statistical properties of our environment.

## Acknowledgments

I would like to thank Kim Plunkett and Nick Chater for very helpful comments on this work.

## References

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48, 71-99.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information.. Psychological Review, 63, 81-97.

Sacks, O. (1985). The Man Who Mistook His Wife for a Hat, Picador: Reading.