

What was that again? A mathematical model of language incorporates the need for repetition

August 29 2011

As politicians know, repetition is often key to getting your message across. Now a former physicist studying linguistics at the Polish Academy of Sciences has taken this intuitive concept and incorporated it into a mathematical model of human communication.

In a paper in the AIP's journal [Chaos](#), Łukasz Dębowski mathematically explores the idea that as humans we often repeat ourselves in an effort to get the story to stick. Using statistical observations about the frequency and patterns of word choice in natural language, Dębowski develops a model that shows repetitive patterns emerging in large chunks of speech.

Previous researchers have noted that long texts have more entropy, or uncertainty, than very brief statements. This tendency to higher entropy would seem to suggest that only through brevity could humans hope to build understanding – uttering short sentences that won't confuse listeners with too much information. But as long texts continue to get longer, the increase in the entropy starts to level off. Dębowski connects this power-law growth of entropy to a similar power-law growth in the number of distinct words used in a text. The two concepts – entropy and vocabulary size – can be related by the idea that humans describe a random world, but in a highly repetitive way.

Dębowski shows this by examining a block of text as a dynamic system that moves from randomness toward order through a series of repetitive

steps. He theorizes that if a text describes a given number of independent facts in a repetitive way then it must contain at least the same number of distinct words that occur in a related repetitive fashion. What this reveals is that language may be viewed as a system that fights a natural increase in [entropy](#) by slowly constructing a framework of repetitive words that enable humans to better grasp its meaning. For now the research is theoretical, but future work could experimentally test how closely it describes real texts, and maybe even candidates' stump speeches.

More information: "Excess entropy in natural language: present state and perspectives" by Lukasz Dębowski is accepted for publication in *Chaos: An Interdisciplinary Journal of Nonlinear Science*.

Provided by American Institute of Physics

Citation: What was that again? A mathematical model of language incorporates the need for repetition (2011, August 29) retrieved 10 April 2024 from <https://phys.org/news/2011-08-mathematical-language-incorporates-repetition.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.