Ferrer i Cancho, R. (2008). Information theory. In: The Cambridge encyclopedia of the language sciences, Colm Hogan, P. (ed.). Cambridge University Press, in press.

INFORMATION THEORY

Information theory is a general purpose and abstract theory for the study of *COMMUNICATION*. Standard information theory was funded by Claude Shannon (1948) and is based on a communication framework in which: (a) the sender must transform a message into a code and send it through a channel to the receiver (b) the receiver must obtain a message from the received code. Communication is successful if the message of the sender is the same as the message obtained by the receiver. For instance, a speaker utters a word for a certain meaning and then hearer must infer the meaning that the speaker had on mind. Noise can alter the code when traveling from the sender to the receiver through the channel (for instance, the speaker produces “Paul” but the “hearer” understands “ball”).

While mainstream linguistics is focused on human language, information theory has been applied to many other contexts, such as the communication systems of other species (McCowan *et al.* 1999, Suzuki *et al.* 2006), genetic information storage in the DNA (Li and Kaneko 1992, Naranan and Balasubrahmanyan 2000) and artificial systems such as computers and other electronic devices (Cover and Thomas 1991).

Information theory has myriad applications even within the domain of the language sciences. We can only give some examples. First, it provides powerful measures in *PSYCHOLINGUISTICS* for measuring the cognitive cost of (a) processing a word (McDonald and Shillcock 2001), (b) an inflectional paradigm (Moscoso del Prado Martín 2004) or (c) the whole *MENTAL LEXICON* (Ferrer i Cancho 2006). Second, information

theory allows one to explain the actual properties of human language. For instance, the tendency of *WORDS* to shorten as the frequency of the word increases can be interpreted as increasing the speed of the information transmitted (e.g., number of messages per second) by assigning shorter codes to most frequent codes. Another well-known property of human language is Zipf's law for word frequencies, one of the most famous *LAWS OF LANGUAGE*. It has been argued this law could be an optimal solution for maximizing the information transmitted when the mean length of words is constrained (Mandelbrot 1966) or maximizing the success of communication while the cognitive cost of using words is minimized (Ferrer i Cancho 2006). Third, information theory has shed light on the *EVOLUTION OF LANGUAGE*. It has been hypothesized that the presence of noise in the communication channel could have favoured the emergence of *SYNTAX* in our ancestors (Nowak and Krakauer 1999), which turns out to be a reformulation of fundamental results from standard information theory (Plotkin and Nowak 2000). Finally, information theory offers an objective framework for studying the differences between *ANIMAL COMMUNICATION AND HUMAN LANGUAGE*. It is well-known that the occurrence of a certain word depends on distant words within the same sequence, e.g., a text, in human language (Montemurro and Puri 2002, Alvarez-Lacalle *et al.* 2006) and information theory studies on other species provide evidence that *LONG-DISTANCE DEPENDENCIES* are not uniquely human (Suzuki *et al.* 2006, Ferrer i Cancho and Lusseau 2006). Furthermore, research on humpback whale songs (Suzuki *et al.* 2006) question Hauser's *et al.*'s (2002) conjecture that only humans employ *RECURSION* to structure sequences.

--Ramon Ferrer i Cancho

Works Cited and Suggestions for Further Reading

Alvarez-Lacalle, Enric, Beate Dorow, Jean Pierrer Eckmann and Elisha Moses 2006. "Hierarchical Structures Induce Long-range Dynamical Correlations in Written Texts". *Proceedings of the National Academy of Sciences of USA* 103.21: 7956-7961.

Cover, Thomas M. and Thomas, Joy A. 1991. *Elements of Information Theory*. New York: Wiley.

Ferrer i Cancho, Ramon 2006. "On the Universality of Zipf's law for word Frequencies". In *Exact methods in the study of language and text. To honor Gabriel Altmann.* ed. Peter Grzybek and Reinhard Köhler, 131-140. Berlin: Gruyter. It discusses the problems of classic models for explaining Zipf's law for word frequencies.

Ferrer i Cancho, Ramon and David Lusseau 2006 "Long-term Correlations in the Surface Behavior of Dolphins." *Europhysics Letters* 14.6: 1095-1101.

Hauser, Marc, Noam Chomsky, W. Tecumseh Fitch. 2002. "The Faculty of Language: what is it, who has it and how did it Evolve?" *Science* 298.5598: 1569-1579.

Li, Wentian and Kunihiko Kaneko. 1992 "Long-range Correlations and Partial $1/f^{\alpha}$ Spectrum in a Noncoding DNA Sequence." *Europhysics Letters* 17.7: 655-660.

Mandelbrot, Benoit. 1966. "Information Theory and Psycholinguistics: a Theory of Word Frequencies". In *Readings in Mathematical Social Sciences* ed. Lazarsfield, P.F., Henry, N.W. 151-168. Cambridge: MIT Press.

McCowan, Benoit, Sean F. Hanser and Laurance R. Doyle. 1999. “Quantitative Tools for Comparing Animal Communication Systems: Information Theory Applied to Bottlenose Dolphin Whistle Repertoires.” *Animal Behavior* 57.2: 409-419.

McDonald, Scott A. and Richard Shillcock. 2001. “Rethinking the Word Frequency Effect. The Neglected Role of Distributional Information in Lexical Processing.” *Language and Speech* 44.3: 295-323.

Montemurro, Marcelo and Pedro A. Pury. 2002. “Long-range Fractal Correlations in Literary Corpora.” *Fractals* 10.4: 451-461.

Moscoso del Prado Martín, Fermín, Alexandar Kostic, and Harald R. Baayen 2004. “Putting the Bits Together: an Information Theoretical Perspective on Morphological Processing.” *Cognition* 94.1: 1-18.

Naranan, Sundaresan and Vriddhachalam K. Balasubrahmanyan. 1998. "Models for Power Law Relations in Linguistics and Information Science". Journal of Quantitative Linguistics 5.3: 35-61. A summary of many optimization models of Zipf’s law for word frequencies that are based on information theory.

Naranan, Sundaresan and Vriddhachalam K. Balasubrahmanyan 2000. “Information Theory and Algorithmic Complexity: Applications to Linguistic Discourses and DNA Sequences as Complex Systems. Part I: Efficiency of the Genetic Code of DNA.” Journal of Quantitative Linguistics 7.2: 129–151.

Nowak, Martin A. and David Krakauer. 1999. “The Evolution of Language”. *Proceedings of the National Academy of Sciences USA* 96.14: 8028-8033.

Plotkin, Joshua, and Martin A. Nowak. 2000. “Language Evolution and Information Theory.” *Journal of Theoretical Biology* 205.1: 147-159.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. Bell Systems Technical Journal 27: 379-423,623-656.

Suzuki, Ryuji, John R. Buck and Peter L. Tyack. 2006. “Information Entropy of Humpback Whale Songs.” *Journal of the Acoustical Society of America* 119.3: 1849-1866.

Zipf, George Kingsley 1935. *The Psycho-biology of Language*. Boston: Houghton Mifflin.