

Towards a Philosophical Grounding of Universal Statistical Properties Underlying Human Language: Mirroring Nature and Rule-Following

Kumiko Tanaka-Ishii

Research Center for Advanced Science and Technology

kumiko@rcast.u-tokyo.ac.jp

May 28, 2018

Big data has revealed new understandings of natural language, namely, that human language universally follows certain statistical properties. These properties appear as power laws, in which the logarithms of two statistical variables are related linearly and the phenomenon is only observable with a certain minimum amount of data. Only a few such power laws are known, but they hold for texts without exception and even beyond texts: daily utterances, including those of child-directed speech, music pieces, and programming language sources all exhibit the same properties, such as shown in (Kobayashi and Tanaka-Ishii 2018). Various cases of human semiosis therefore seem to follow some kind of identical characteristic, or rule.

Curiously, we never conduct linguistic actions intending to produce such properties, yet the accumulation of such actions inevitably leads to these properties. Therefore, a reason these properties hold must exist. In the scientific domain, the reason remains unrevealed and will likely remain so without a significant leap in neuroscience. Hence, the subject for now remains a philosophical discussion. This talk attempts to comprehend what these universal properties might signify, from two different perspectives.

The first perspective is the classical viewpoint of “mirroring nature.” Power laws have been abundantly reported for natural and social systems, which are studied as complex systems in the physics domain. If we live within such a system, then our representational system of language must be able to accommodate the target structure. In the talk, some scientific explanations (Mandelbrot 1952; Heidelberger and Klohr 2004) of the universal properties, which take such a *mirroring* approach, will be summarized.

The mirroring approach does not, however, provide any understanding on how exactly universal properties emerge. As mentioned before, each such property is global and observable with a certain amount of data, but this occurs through accumulation of individual acts. In other words, we cannot be consciously generating these properties; instead, we simply use language word after word, possibly following some rules as Wittgenstein (1953) demonstrated, and the consequence inevitably leads to universal statistical properties. If the origin of these properties lies in individual acts, then what stipulates every act of rule following? Or, what is the nature of the rules?

The origin of universal properties is thus considered from this second perspective of

rule following. Specifically, recent mathematical formulations to explain the emergence of statistical properties by generative processes have a strong relation with philosophical debates on rule following (McGinn 1984; Kripke 1982; Miller and Wright 2014; Kusch 2006). A generative process stipulates the following event (word) according to some rule, given a past sequence of generated events. The emphasis of generative processes differs, however, between mathematics and philosophy. In philosophy, the central issue is how a repeated event possibly possesses the same meaning as the past events, whereas in mathematics, the main issue lies in the rules for introducing new events. Regarding this introduction of new events, there is a debate by McGinn (1984) in relation to the passage (Wittgenstein 1953)[199], but the argument within the themes of rule following and private language is deemed to remain few.

In the talk, I will discuss how a new word is initiated within private language in order to present a new meaning. Such introduction of new words within a context of repeated words gradually extends the communication sphere to form a complex system. The complex structure that emerges through the accumulation of such rule-following acts is the source of the statistical laws of language, forming one basis of our language game.

References

- Heidelberger, M. and Klohr, C. (2004). *Nature from Within: Gustav Theodor Fechner and His Psychophysical Worldview*. University of Pittsburg Press.
- Kobayashi, T. and Tanaka-Ishii, K. (2018). Taylor’s law for human linguistic sequences. *In Proceedings of Annual Conference on Computational Linguistics*. arXiv:1804.07893.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press. Reprint in 1984.
- Kusch, M. (2006). *A sceptical guide to meaning and rules: Defending Kripke’s Wittgenstein*. McGill Queens University Press.
- Mandelbrot, B. (1952). An informational theory of the statistical structure of language. *Proceedings of SYmposium of Applications of Communication theory*, pages 486–500.
- McGinn, C. (1984). *Wittgenstein on Meaning: An Interpretation and Evaluation*. Basil Blackwell.
- Miller, A. and Wright, C. (2014). *Rule-Following and Meaning*. Routledge.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Willey-Blackwell. 4th reprint 2009.