Iconicity Through the Lens of Information Theory

Introduction

In both functional-cognitive linguistics and semiotics, the concept of iconicity revolves around the perceived similarity or analogy between the form of a linguistic sign and its associated meaning, standing in contrast to the principle of arbitrariness. Harnessing insights from information theory, our study seeks to unravel how transitional probabilities contribute to the evocativeness of English words and influence the iconicity ratings assigned to them. Through this information-theoretic lens, we aim to illuminate the cognitive mechanisms at play in the intricate relationship between language and iconicity, presenting new perspectives on the profound interplay between linguistic form and conceptual meaning.

In the development of information theory (Shannon, 1948), two fundamental concepts, surprisal and entropy, emerged as essential measures in quantifying uncertainty and information content within a communication system. Surprisal refers to the amount of surprise or unexpectedness associated with the occurrence of an event. Mathematically, surprisal (S) is defined as the negative logarithm of the probability (P) of the event square, and it can be expressed as: $S = -log_2(P)$. Entropy, on the other hand, is a measure of the average information content or uncertainty within a set of events. It represents the average surprisal associated with all possible outcomes, reflecting the level of disorder or randomness in the system. Mathematically, entropy (H) is calculated as the sum of the surprisal of each event (S) multiplied by its probability (P): $H = \Sigma(P_i * -log_2(P_i))$.

Surprisal and entropy have been shown to influence speech production (Cohen Priva, 2015; Hume & Mailhot, 2013) and perception (Kilpatrick et al., 2021; Whang, 2021). These studies have suggested that there are limitations on how much information is expressed over time in human languages. Pellegrino et al., (2011) conducted a cross-linguistic analysis of 17 languages to investigate the information rate of speech. They measured the average information density in speech, expressed in bits per second, and found a remarkable convergence in the average information rate (approx. 39 bit/sec). They propose that this convergence is potentially due to universal cognitive and perceptual limitations of humans.

Building on this finding, we hypothesize that words carrying more transitional probability information may carry less iconicity information and vice versa. In the present study, this might manifest in one of two ways:

H1) Words with lower average surprisal will have higher iconicity ratings.

H2) Words with higher average entropy will have higher iconicity ratings.

Method

Data and scripts can be found at the following OSF repository: https://osf.io/qh8vs/

Data were taken from a corpus of 14,000 English words with iconicity ratings given by American English speakers (Winter et al., 2023). Words were transcribed into American English using the CMU pronouncing dictionary (Weide, 2015). Any word in the corpus that did not find a match in the CMU pronouncing dictionary was discarded resulting in 13,888 words. Word length was calculated by totaling the number of phonemes in each word. Bigram surprisal and entropy were calculated for each word. Average surprisal was calculated by totaling the surprisal values for each bigram in each word and dividing it by the number of phonemes less one (total/N-1). Average entropy was calculated by totaling the entropy scores for each phoneme in the word and dividing it by the number of phonemes (total/N).

Results

Two simple linear regression models were employed to investigate the relationship between surprisal, entropy, and iconicity ratings of words. The results from the initial models indicated statistically significant associations. The surprisal model showed a significant negative association ($F(1, 13880) = 117.8$, $p < 0.001$), while the entropy model demonstrated a significant positive association ($F(1, 13886) = 138.3$, $p < 0.001$). These findings suggest that words with higher iconicity ratings carry less transitional probability information (surprisal) and are more predictable (entropy). Additional simple linear regression models were used to examine the relationship of word length with all other variables. These showed a significant positive correlation between length and entropy ($p < 0.001$) and significant negative correlations with both surprisal ($p < 0.001$) and iconicity ratings ($p < 0.001$); i.e., longer words carry less iconicity information and are more transitionally predictable.

Subsequently, a more comprehensive analysis was conducted by introducing word length as an additional independent variable in multiple linear regression models. These models revealed the interactive effects of average entropy and average surprisal with word length on the dependent variable, iconicity ratings. The interactions were assessed through the inclusion of interaction terms (average surprisal/length and average entropy/length) in the models. The interaction between average surprisal and length on rating was significant ($t = 5.345$, $p < 0.001$) as was the interaction between average entropy and length on rating ($t = -2.799$, $p = 0.005$); i.e., the effects of surprisal and entropy on ratings vary depending on word length.

Discussion

Although this is only a preliminary study, the findings suggest that words with higher iconicity ratings tend to carry less transitional probability information, particularly when considering the interaction with word length. The results emphasize the complexity of the relationship between iconicity and transitional probability, warranting further investigation into the mechanisms underlying this association. As shown by Pellegrino et al. (2011), languages are limited in how much transitional probability information they can express per second, likely due to cognitive limitations in language users. These results suggest that incorporating iconicity information in language processing may also impose cognitive burdens on language users, potentially influencing language production, comprehension, and change. Given the intriguing interplay between iconicity and transitional probability, future research could explore the cognitive mechanisms and trade-offs involved, thereby contributing to a more comprehensive understanding of language structure and processing.

References

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. Laboratory phonology, 6(2), 243-278.

Hume, E., & Mailhot, F. (2013). The role of entropy and surprisal in phonologization and language change. In A. C. L. Yu (Ed.), Origins of sound change: Approaches to phonologization (pp. 29–47). Oxford University Press.

Kilpatrick, A., Kawahara, S., Bundgaard-Nielsen, R., Baker, B., & Fletcher, J. (2021). Japanese perceptual epenthesis is modulated by transitional probability. Language and Speech, 64(1), 203-223.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.

Weide, R. L. (1998). The Carnegie Mellon pronouncing dictionary. [Computer software]. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Whang, J. (2021). Multiple sources of surprisal affect illusory vowel epenthesis. Frontiers in Psychology, 12, 677571.

Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. (2023). Iconicity ratings for 14,000+ English words. Behavior research methods, 1-16.