Efficient use of ambiguity in an early writing system: Evidence from Sumerian cuneiform

Noah Hermalin[1] **(nmhermalin@berkeley.edu)**
Terry Regier[1,2] **(terry.regier@berkeley.edu)**

[1]Department of Linguistics and [2]Cognitive Science Program
University of California, Berkeley, CA 94720 USA

Abstract

Ambiguity has often been viewed as a hindrance to communication. In contrast, Piantadosi et al. (2012) argued that ambiguity may be useful in that it allows communication to be efficient, and they found support for this argument in the spoken forms of modern English, Dutch, and German. The historical origins of this phenomenon cannot be probed in the case of spoken language, but they can for written language, as it leaves an enduring trace. Here, we explore ambiguity and efficiency in one of the earliest known written forms of language: Sumerian cuneiform. Sumerian cuneiform exhibits extensive ambiguity, and for that reason it has been considered to be poorly suited for communication. We find, however, that ambiguity in Sumerian cuneiform supports efficient communication, mirroring the earlier findings for spoken English, Dutch, and German. Thus, the early stages of human writing exhibit evidence suggesting pressure for communicative efficiency.

Keywords: efficient communication; ambiguity; writing systems; cuneiform; information theory

Introduction

Ambiguity in language is often considered to be communicatively disadvantageous, because it can make a speaker's intention unclear to a listener. However, it has been argued (Zipf, 1949) that a certain amount of ambiguity in language is inevitable given the competing needs of speakers and listeners. Piantadosi, Tily, and Gibson (2012) pursued this idea further, and argued that ambiguity may be useful in that it can support efficient communication. They showed empirically that patterns of ambiguity in the spoken forms of modern English, Dutch, and German are consistent with pressure for efficiency in communication.

Given this finding, it is natural to wonder about the historical origins of this phenomenon. How quickly do languages come to exhibit efficient use of ambiguity? Was this phenomenon present near the beginning of language use? Such questions are unanswerable for the spoken form of language, which leaves no lasting trace — but they can be addressed with respect to written language, which does leave such a trace.

Sumerian cuneiform is one of the earliest known writing systems, and is one of the four 'pristine' writing systems of the world, meaning that its origins are not traceable to borrowing or influence from any previously existing writing system (Woods, 2015a). It is also known to be highly ambiguous, such that a given character often has numerous distinct semantic and/or phonological values (Cooper, 1996). Additionally, the distribution of meanings across forms in written Sumerian was not simply a straightforward reflection of spoken Sumerian; this means that any finding of efficiency with respect to the writing system cannot be dismissed as entirely derivative of the corresponding spoken language. Finally, Sumerian is unrelated to the languages studied earlier by Piantadosi et al., which are closely related to each other. For these reasons, Sumerian cuneiform suggests itself as a natural case study for probing the historical origins of the efficient use of ambiguity, in the accessible case of written language.

In what follows, we first provide a brief introduction to Sumerian cuneiform, and its relevance to the question of ambiguity and efficiency. We then restate the argument and results of Piantadosi et al. on modern spoken languages. Then, in three studies, we apply the logic and methods of Piantadosi et al. to the problem of assessing efficiency in Sumerian cuneiform. We find that ambiguity in Sumerian cuneiform bears the same signatures of efficiency as were found in modern spoken languages. We conclude that pressure for efficient communication may have been present near the earliest stages of human writing, and we discuss the implications of this conclusion.

Sumerian cuneiform

Cuneiform writing developed in southern Mesopotamia throughout the 4th millennium BC; first used for linguistic writing by the 31st century, the system survived roughly three thousand years, over which it was adapted into various languages of the Middle East (Veldhuis, 2012). The first language for which cuneiform was used was most likely Sumerian (Veldhuis, 2012), an agglutinative language with mild nominal morphology (case-marking suffixes) and rich verbal morphology, including a plethora of tense-aspect-mood and agreement affixes (Michalowski, 2004).

Cuneiform tablets compartmentalized text into columns, which were further divided into lines/cells, somewhat similar in layout to a modern-day spreadsheet; smaller items would only have one column (see Figure 1 for an example). The amount of information contained within a cell of a text had some degree of variation, but was at least at the level of a word and typically at the level of a phrase. Earlier scribal practice was not always concerned with preserving a consistent linear order of characters within a cell. By c. 2400 BC, however, scribes adhered to fairly strict and consistent linearity in spellings (Michalowski, 2004).

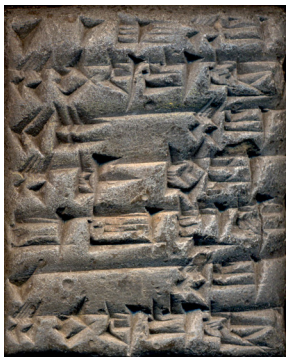


Figure 1: A sample text in Sumerian cuneiform. Since the text is small, it only has one column, which is divided into seven lines. Image from the Cuneiform Digital Library Initiative (2016), CDLI #102525. Image reprinted with permission of Robert K. Englund.

Spelling	Transliteration	Meaning
	paŋ	'breathe'
	asag	'demon'
	pa	'branch'
	ugula	'overseer'
	sag	'beat'
	rig	'boil down'
	rig	'donate'
	ensi	'governor/ruler'
	maškim	'administrator'
	munsub	'shepherd$_1$'
	sipad	'shepherd$_2$'

Table 1: Non-exhaustive list of words that contain the character in their spelling.

Written Sumerian was primarily logographic: the level of linguistic representation for a given graphical unit would usually be the morpheme (or some sub-morphemic, non-phonemic unit of information), although it also made use of phonography to some degree, with characters sometimes mapping more directly to (strings of) sounds, usually at the level of the syllable (Civil, 1973). A major feature of the system was its extensive use of ambiguity: any given character could have numerous distinct semantic and/or phonological values (Cooper, 1996). A non-exhaustive list of words containing the character can be found in Table 1; this list serves as an example of how a single character may occur in the spellings of words which do not all share semantic or phonological information. The table also demonstrates the lack of strict isomorphism between written form and corresponding spoken form, either in terms of phonemes, syllables, or morphemes. This point is important for our purposes because it means that if written Sumerian bears signs of efficiency, that efficiency cannot have been entirely inherited from spoken Sumerian.

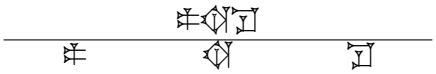
Two open questions concerning efficiency and ambiguity emerge from this overview. First, and most centrally: is the ambiguity of Sumerian cuneiform communicatively harmful, as might be expected given its extensiveness — or is it instead consistent with pressure for efficiency in communication? Second: is the shift to greater linearity in writing attributable to pressure for efficiency? We pursue these questions below.

The argument of Piantadosi et al. (2012)

To address these questions, we draw on the logic and methods of an earlier study that focused on modern spoken languages. Piantadosi et al. (2012) argued that "ambiguity is a functional property of language that allows for greater communicative efficiency" (p. 280). Their argument coheres naturally with a classic functionalist view that seeks to explain language structure and use in terms of efficient communication, and a growing body of recent research that pursues that idea with respect to various aspects of language (e.g. Aylett & Turk, 2004; Ferrer i Cancho & Solé, 2003; Piantadosi, Tily, & Gibson, 2011; Fedzechkina, Jaeger, & Newport, 2012; Kirby, Tamariz, Cornish, & Smith, 2015; Kemp, Xu, & Regier, 2018).

Piantadosi et al. (2012) pursued this argument as follows. First, they argued that context has the potential to resolve ambiguity. The communicative problem posed by ambiguity is that of the listener's (or reader's) uncertainty about the meaning of a given form, and they engaged this problem in information-theoretic terms, casting uncertainty as entropy. They noted that if context is informative about meaning, context will necessarily reduce uncertainty (entropy) about meaning. This means that context has the potential to alleviate the problem posed by ambiguity: a form that may be highly ambiguous in isolation may be much clearer when considered in context. A central assumption of their paper is that context is in fact informative about meaning, and therefore does help to disambiguate.

Piantadosi et al. then pursued the hypothesis that ambiguity in language is deployed in a manner that increases efficiency. The core idea is that if ambiguity is resolved by context, forms are free to take on multiple meanings — and the efficient way to do this would be to preferentially re-use forms that are low-cost, so as to minimize overall cost, or effort (Zipf, 1949). Forms may be low-cost in various ways: they may be short or otherwise simple; they may be frequent and therefore processed more quickly, and so on. Their paper considered several measures of form cost, and asked to what extent each predicts ambiguity of form. Specifically, using data on the spoken forms of German, Dutch, and English, they conducted quasi-Poisson regressions to establish the relationship between various count measures of form ambiguity and three properties of form cost: length, frequency (as negative log probability), and phonotactic surprisal. They found that in general, greater ambiguity was predicted by lower form cost (with the possible exception of phonotactic surprisal). Thus, shorter and more frequent forms were more ambiguous in



(RULER, 1) (RULER, 2) (RULER, 3)

Table 2: Example morpheme, meaning 'ruler'. The top row shows how this morpheme would be spelled in characters in the original text. The bottom two rows show the characters that appear in the spelling of this morpheme, each paired with the value of that character with respect to this morpheme, as defined in Equation 1.

German, Dutch, and English — consistent with the expectation that low-cost forms are preferentially re-used, as predicted by pressure for efficiency.

The present studies

We applied an analogous line of investigation to the question of ambiguity in Sumerian cuneiform.[1] Ambiguity arises when a form has more than one value, or symbolic function. Thus, to explore ambiguity in Sumerian cuneiform, we need to specify the relevant unit of form, and the corresponding values. It is natural to take the character as the relevant unit of form in Sumerian cuneiform, as it is characters that are often considered to be ambiguous. And given that characters are not defined either purely semantically or purely phonologically, but are used to specify morphemes, it is natural to define the values of a character in terms of that character's role in identifying morphemes, i.e. the character's role in spelling morphemes. A morpheme can have more than one spelling, so we first define the spellings $S(m)$ of a morpheme m to be the set of character strings that spell out that morpheme in Sumerian cuneiform. We then define the values *V(x)* of a character x as:

$$V(x) = \{\ (m,i)\ \mid\ x \text{ is the } i^{\text{th}} \text{ character in } s \in S(m)\ \} \quad (1)$$

That is, the values of character x are the set of all (morpheme, index) pairs (m,i) such that x is the i^{th} character in one of the spellings s of morpheme m. For example, the values of the character 𒁹 include the pairs (RULER, 1), (BRANCH, 1), and (DEMON, 2), among others. Table 2 illustrates a spelling of a specific morpheme, and the determination of character values from that spelling.[2]

Data

The data we used were from ORACC, the Open Richly Annotated Cuneiform Corpus (Tinney & Robson, 2014), an open-access corpus of cuneiform texts which is, to the best of our knowledge, the largest open-access corpus for Sumerian texts that has POS tagging and morphological annotation. Specifically, we used the texts in the Ur III Administrative Documents corpus within ORACC; this corpus is roughly 5.5 million cuneiform characters in length, and it consists of various administrative and transactional documents from the Ur III period (c. 2112-2004 BC). This corpus was chosen because it is the largest single-genre morphologically annotated corpus of third millennium Sumerian texts.

Substantial parts of the corpus had to be discarded. We omitted tokens that were damaged or for which the reading was unknown. In addition, most proper nouns had to be omitted.[3] The resulting cleaned data had roughly 3.3 million character tokens. We refer to this cleaned corpus as the 'dataset'.

Overview of the present studies

We conducted three studies to test whether ambiguity in Sumerian cuneiform is consistent with pressure for efficient communication. Piantadosi et al. (2012) assumed that much ambiguity could be resolved by context; we wished to test this question directly, so Study 1 asks to what extent context resolves ambiguity in Sumerian cuneiform. Study 2 asks whether context disambiguates more effectively in Sumerian cuneiform than it does in a number of plausible hypothetical variants of it; in so doing, this study explores whether increasing linearity in Sumerian writing may have resulted from pressure for efficiency. Finally, Study 3 applies the analyses of Piantadosi et al. (2012) to Sumerian cuneiform, to determine whether the signatures of efficiency they found in modern spoken languages are also found in cuneiform.

Study 1: Does context disambiguate?

To what extent does context resolve ambiguity in Sumerian cuneiform? We considered a simple version of this general question. We first determined the uncertainty concerning which value a character has when the reader knows only the current character (unigram condition). We then compared this to the uncertainty when the reader knows not just the current character but also the preceding character (bigram condition).

We took uncertainty concerning character values to be the conditional entropy of values V conditioned on context C:

$$H(V|C) = -\sum_{c \in C} P(c) \sum_{v \in V} P(v|c) \log_2 P(v|c) \quad (2)$$

Lower conditional entropy denotes greater certainty concerning character value.

We calculated $H(V|C)$ over the entire dataset, once taking C to be the current character alone (unigram), and once again taking C to be the current and preceding characters together (bigram). The results are shown in the top two lines of Table 3. Conditional entropy in the bigram condition is much lower than in the unigram condition. This demonstrates not only that context disambiguates, but also that just a single

[1] We believe we are the first to treat Sumerian in this way. However Civil (1973) informally explored the possibility of examining Sumerian cuneiform through the lens of information theory.

[2] We also ran all of the analyses using an alternate definition of a character's values: $V(x) = \{\ m\ \mid\ x \text{ is present in } s \in S(m)\ \}$. By this definition, a character x's values are simply the set of morphemes that contain x anywhere in any of their spellings. The results using this definition of character values were qualitatively the same as the results reported here.

[3] Proper nouns had no morphological annotation. Among other problems, this meant that inflectional morphology was not automatically separable from the rest of the word for proper nouns, as it was for other words in the corpus.

Table 3: Conditional entropy $H(V|C)$ of character values V given one character (unigram) vs. two characters (bigram) of text C, on attested and hypothetical data. Study 1: One added character of context results in a sharp decrease in uncertainty in attested data. Study 2: context disambiguates more effectively in attested Sumerian cuneiform than it does in some hypothetical variants of it. Value for WLSS is the average $\pm$ 1 SD, over 500 systems.

Study	Condition	$H(V\|C)$
1	Unigram, attested data	1.5281
1	Bigram, attested data	0.4584
2	Bigram, BWS	0.4719
2	Bigram, WLSS	0.9796 ($\pm$ 0.0004)

additional preceding character of context suffices to substantially reduce uncertainty. Since much more context than this would be available to readers, it is reasonable to expect that a competent reader of Sumerian would be able to infer with high certainty which value a given character was intended to have, in context. We conclude from this finding that context does effectively disambiguate in Sumerian cuneiform.

Study 2: Comparison with hypothetical systems

Given that context disambiguates in Sumerian cuneiform, we ask the follow-up question of whether plausible hypothetical variants of the system exhibit better, worse, or comparable results. Study 2 tested whether the consistency of spellings and strict linearity of Sumerian cuneiform demonstrate advantages over hypothetical competitors with regards to certainty of decoding character values in context.

We considered two hypothetical variants of Sumerian cuneiform. The first variant is 'backwards Sumerian' (BWS): this is a hypothetical variant of Sumerian in which the entire corpus is spelled backwards. Effectively this means that when considering a character in context, we take as context what would have been the following character in actual Sumerian, rather than the preceding character as in Study 1. The other hypothetical variant is 'within-line shuffled Sumerian' (WLSS): this is a system that is derived from our Sumerian cuneiform dataset by randomly shuffling the order of characters within a line. In this case, a neighboring character taken as context could be any other character within the same line in the original dataset. The latter hypothetical variant is motivated to some extent by actual scribal practices in earlier periods, in which characters were not always arranged in linear order. It is known that written Sumerian shifted towards more consistent linearity over time (Michalowski, 2004), and these hypothetical variants allow us to test the hypothesis that the greater linearity that we see in Ur III written Sumerian (the period of our dataset) may have aided disambiguation.

We first calculated $H(V|C)$ over the BWS datset. We then generated 500 WLSS datasets by randomly reordering characters and their respective values within each line, and calculated $H(V|C)$ over each resulting WLSS dataset. We considered only the bigram condition (in which C is the current character together with an immediately preceding character), because the unigram condition would yield identical results in the attested and hypothetical systems.

The results are shown in Table 3. Bigram conditional entropy is very slightly higher for BWS than it is for the attested system; thus, following context may serve as a marginally weaker disambiguator than preceding context, but the difference is small. Bigram conditional entropy is substantially higher for the WLSS systems than it is for the attested system, demonstrating that consistent linearity of spelling does confer an advantage on an ambiguous, logographic system such as Ur III written Sumerian, at least with respect to determining a given character's value based on immediately neighboring context. These results elaborate those of Study 1, and show that context disambiguates more effectively in Sumerian cuneiform than it does in at least some hypothetical variants of that system.

Study 3: Is ambiguity used efficiently?

We have seen that the ambiguity of written Sumerian is much reduced by contextual information, and that this is more true of actual Sumerian than it is of some possible variants of it. This sets the stage for a question directly parallel to that posed by Piantadosi et al.: given that context disambiguates, do languages use ambiguity efficiently, by reusing low-cost (simple, frequent) forms for a large number of meanings, thereby reducing system-wide cognitive costs?

We addressed this question in a manner that mirrors that of Piantadosi et al.: by asking whether the number of values associated with a specific character was predicted by the character's frequency of occurrence in the dataset, and by its simplicity.[4] Our measure of complexity (the opposite of simplicity) for a cuneiform character was stroke count: the number of strokes or wedges required to produce the canonical form of the character. For example, the character 𒐼 has 3 strokes. Stroke counts were coded manually by the first author based on forms in the Electronic Pennsylvania Sumerian Dictionary Project (ePSD; Tinney, 2009), an open-access online dictionary. Following Piantadosi et al., we transformed character frequency to negative log (unigram) probability, using add-one smoothing so that no character had frequency zero.

Figure 2 plots the number of values a character has (its character valence, $|V(x)|$, which is the size of the set $V(x)$), as a function of negative log probability based on unigram frequency, and as a function of stroke count. In both cases it appears qualitatively that lower-cost (more frequent, simpler) characters tend to have more values, consistent with pressure for efficiency.

To probe this pattern quantitatively, we conducted a quasi-Poisson regression to predict the number of values $|V(x)|$ associated with each character x, from that character's negative log probability and stroke count. We standardized the two pre-

[4]The third predictor considered by Piantadosi et al., phonotactic surprisal, is not applicable to Sumerian cuneiform.

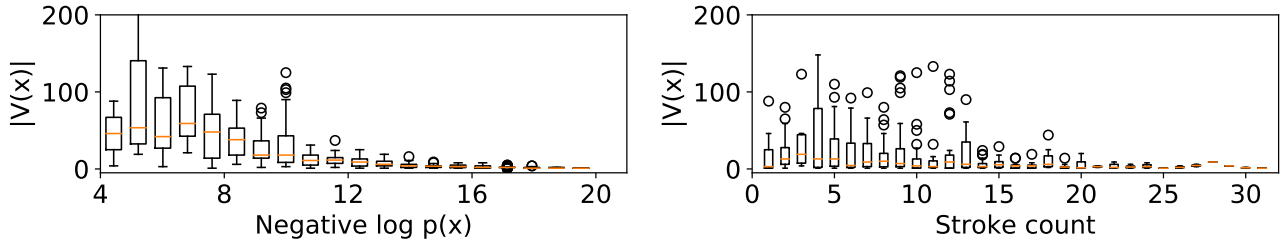


Figure 2: Box plots showing character valence $|V(x)|$ as a function of (left panel) negative log probability based on character frequency, and (right panel) character stroke count. Boxes extend from lower to upper quartiles; orange lines denote median; whiskers extend to 1.5 times the interquartile range beyond the lower and upper quartiles; all data points not in this range are treated as outliers and shown as empty circles. Both panels suggest that low-cost forms are preferentially re-used: more frequent (lower $-\log p(x)$) characters tend to have more values, and simpler (lower stroke count) characters tend to have more values.

dictor variables: for each variable, we subtracted the mean of that variable and divided by one standard deviation. This regression revealed significant effects both of negative log probability and of stroke count. Higher negative log probability (lower frequency) was negatively associated with number of values ($\beta = -1.249, t = -19.792, p < 0.001$), meaning that higher frequency characters were associated with more values. To make this outcome concrete, consider that the most frequent 180 characters, which make up only 27% of the total number of character types in the dataset, bear 66% of all values. Thus, a reader only needs intimate familiarity with a modest number of characters in order to be fairly literate. Higher stroke count was also negatively associated with number of values ($\beta = -0.127, t = -2.072, p < 0.05$), meaning that simpler characters (those with fewer strokes) were associated with more values.

Thus, characters with more values tend to be both more frequent and graphically simpler, as predicted by the hypothesis of efficiency: Sumerian cuneiform exhibits preferential re-use of low-cost material.

Discussion

The traceable origins and early years of written language offer a unique window into the role that pressure for efficient communication can play in shaping linguistic systems. For this reason, the present study has explored efficiency in one of the earliest known writing systems: Sumerian cuneiform, the written form of the Sumerian language.

We have seen that written Sumerian bears signs that are consistent with the hypothesis of pressure for efficient communication. Despite the high degree of ambiguity in written Sumerian, we have seen that a reader would only need a small amount of additional context to be able to decode a character's value with high certainty (Study 1). We have also seen that a comparison with hypothetical alternate systems which deviate from canonical linearity suggests that the system may have gravitated towards a more consistent linearity of spelling in a way that allowed for increased certainty of decoding (Study 2). Finally, we have seen that since context serves to reliably disambiguate character values, the system was able to use a single given form for several different values without sacrificing system informativeness — and that it appears to have done so in an efficient manner, preferentially re-using low-cost forms (Study 3). Taken as a whole, this evidence shows that written Sumerian was not an inefficient system.

Several general implications can be drawn from this observation. One of these concerns efficiency in writing systems generally. While factors such as medium (e.g. Woods, 2015b) and societal pressures (e.g. Veldhuis, 2012) are undoubtedly relevant to the development of a written language, our results demonstrate that pressures of communicative efficiency have acted on written systems since the earlier days of writing itself. Despite the relative disconnect between written Sumerian and its corresponding spoken language in terms of how values are distributed across contrastive units, the same signature of efficiency that Piantadosi et al. (2012) observed in three spoken languages in is also found in Ur III written Sumerian. This suggests that pressure for efficient communication is not unique to spoken or signed language, but is present in written language as well — critically, even when the written language does not closely mirror a corresponding spoken language. Thus, communicative efficiency may be viewed as a general principle of linguistic communication independent of medium or modality.

Another potential implication concerns the time course of the presumed cultural evolutionary process that produces efficiency in linguistic systems. The fact that our results were obtained in a linguistic system as young as 1000 years old suggests that these pressures may act upon a system from its inception and guide it toward greater efficiency within a comparatively short period of time. Since our analyses do not include actual data from periods earlier than Ur III we cannot be completely sure that earlier periods would have been less efficient. However, the fact that our hypothetical shuffled system performed poorly relative to the Ur III corpus is at least suggestive that earlier texts, which were analogously less consistent with their linearity, may not have evolved the specific communicatively useful features we have documented for Ur III Sumerian. Settling this question more definitively would require a thorough comparison of efficiency across ear-

lier time periods, tracking the progression of written Sumerian toward the system we have investigated.

In addition to a direct comparison of written Sumerian across earlier timer periods, future work on this topic would benefit from a more thorough consideration of the psycholinguistic evidence regarding recognition, decoding, and storage of graphical units. While we considered stroke count as a measure of visual complexity (which can be detrimental towards character recognition and processing, especially at lower frequencies; see e.g. Tamaoka & Kiyama, 2013), we did not consider other factors such as visual similarity between characters. Finally, future work could usefully consider the consequences of using the same (or similar) characters for phonologically or semantically related morphemes.

Firmer, broader, and more detailed conclusions will have to await the outcome of such possible future research. For now, however, we can conclude on the basis of the evidence we have seen here that one of the earliest known writing systems exhibits patterns of ambiguity that are consistent with pressure for efficient communication.

Acknowledgments

We thank Robert K. Englund for giving us permission to reprint the image shown in Figure 1. This study was supported in part by the Defense Threat Reduction Agency; the content of the study does not necessarily reflect the position or policy of the U.S. government, and no official endorsement should be inferred.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*, 31–56.

Civil, M. (1973). The Sumerian writing system: Some problems. *Orientalia*, *42*, 21–34.

Cooper, J. S. (1996). Sumerian and Akkadian. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 37–72). New York: Oxford University Press.

Cuneiform Digital Library Initiative. (2016). Retrieved from `http://cdli.ucla.edu`

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*, 17897-17902.

Ferrer i Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*, 788–791.

Kemp, C., Xu, Y., & Regier, T. (2018). Semantic typology and efficient communication. *Annual Review of Linguistics*, *4*, 109–128.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87-102.

Michalowski, P. (2004). Sumerian. In R. D. Woodard (Ed.), *The Cambridge encyclopedia of the world's ancient languages* (pp. 19–59). Cambridge: Cambridge University Press.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526-3529.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

Tamaoka, K., & Kiyama, S. (2013). The effects of visual complexity for Japanese kanji processing with high and low frequencies. *Reading and Writing*, *26*, 205–223.

Tinney, S. (2009). *Electronic Pennsylvania Sumerian Dictionary project.* Retrieved from `http://psd.museum.upenn.edu`

Tinney, S., & Robson, E. (2014). *ORACC: The Open Richly Annotated Cuneiform Corpus.* Retrieved from `http://oracc.museum.upenn.edu`

Veldhuis, N. (2012). Cuneiform: Changes and developments. In S. D. Houston (Ed.), *The shape of script: How and why writing systems change* (pp. 3–23). Santa Fe, NM: School of Advanced Research Press.

Woods, C. (2015a). The earliest Mesopotamian writing. In C. Woods, G. Emberling, & E. Teeter (Eds.), *Visible language: Inventions of writing in the ancient Middle East and beyond.* (pp. 33–50). Chicago: The Oriental Institute of the University of Chicago.

Woods, C. (2015b). Visible language: The earliest writing systems. In C. Woods, G. Emberling, & E. Teeter (Eds.), *Visible language: Inventions of writing in the ancient Middle East and beyond.* (pp. 15–25). Chicago: The Oriental Institute of the University of Chicago.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Reading, MA: Addison-Wesley.