The “Universal” Rebus Principle and Phonosemantic Compounding

John Kausch
Western University of Ontario, Canada

Abstract

The general public views hieroglyphs as “picture-writing”. Why does this view persist after Champollion showed that the key to reading hieroglyphs is phonetic? This article argues that this misinterpretation is a historical remnant of the transmission and reception of a single text, the Hieroglyphica of Horapollo. The content of Hieroglyphica is reinterpreted as a conflation of the rebus principle with Egyptian ideas of sound symbolism. The reception of this text in the late Renaissance created the idea of a universal language as an unambiguous pictorial writing system, which has never existed for natural language. Sound
symbolism, or phonosemantics is discussed with a particular eye towards the challenges of creating universal languages, including both the difficulty of representing abstract concepts, and the strain on the memory of the reader. The article discusses a solution to these problems of ambiguity in the construction of an artificial rebus principle. A prototype for a set of pictographic universal concepts mapped to characters in the International Phonetic Alphabet (IPA) is then proposed and discussed. These pictograms are based on a swadesh list that was extended by the very large lexicon PanLex, and make use of both hieroglyphs and emoji as glyphs.

Keywords: phonosemantics, rebus principle, hieroglyphica, research creation

1. Introduction

The employment of the rebus principle defines the boundary between proto-writing and true writing. The general public’s conceptions of logographic writing systems like Egyptian, Chinese, Sumerian or Mayan writing are what linguists would describe as proto-writing. A brief look at the articles published in the media, and even some of the academic literature about emoji which compare them to Egyptian hieroglyphs confirms the presence of this misconception (Alshenqeeti 2016, Reale et al. 2021). This misconception was historically widespread before Champollion’s deciphering of Egyptian hieroglyphs. It had a remarkable influence on figures such as Bacon or Leibniz, who are responsible for popularizing the idea of universal languages. Therefore, it is useful to investigate this history as a way to interrogate the linguistic assumptions on which universal languages are predicated.

It is difficult for moderns to understand the set of social relations and philosophical assumptions which structured the role of writing in the Bronze Age. In many respects the metaphysical orientation of
peoples of the ancient Near East is foreign to contemporary thought. This includes the writing systems with which they wrote. The current social order, especially the Western social order, is based on an underlying assumption of the primacy of alphabetic writing as the default mode for writing systems. Thus, when moderns imagine semantic writing systems, either among non-specialist contemporaries or historically, these assumptions inevitably color the way pictographic or logographic writing is imagined.

Linguists, as educated specialists, have a more sophisticated understanding of the different writing systems which are possible to represent human language, and take a typological approach to model this variation (Borgwaldt & Joyce 2013). Yet, when applied to history, this typological approach fails to capture the complex interplay between word and image which lead to the emergence of writing. The basic contours of this historical problem are the line which is drawn between proto-writing and “true writing”. The theoretical principle which separates the two is the rebus principle, as it is the mechanism by which pictographic signs are transformed into ones with phonological content.

While scientifically useful as a theory, it is an anachronism to assume that the ancient peoples who created ideographic or logographic writing systems would have understood what they were doing in such abstract terms. The theory describes a historical process by which iconic symbols were transformed into symbols which manipulated language: but the status of the rebus principle, and what it meant to those who manipulated these first symbols, is not a mere historical curiosity. When applied to the design principles necessary to the creation of a semantic writing system the rebus principle, and the typology which implicitly underlies it, becomes a theoretical assumption which it is necessary to interrogate.
That this is a problem becomes evident the moment we consider the design problems in creating a universal semantic writing system. There are two primary semiotic issues in the creation of a purely pictographic communication system. The first is the inherent ambiguity of images to represent things and concepts (especially abstract concepts), and the second is the strain on the memory that is produced for the reader if signs are created to represent all visible entities (Ghorbanpour 2022, New 2022). In the discussions of semantic writing systems which have occurred since the Enlightenment, there is a general assumption that such a pictographic system that is visual and unambiguous, or at least readily understood, is possible.

All of the historical analogs we have for semantic writing systems (Egyptian, Sumerian, Chinese and Mayan writing being the most famous exemplars, although there are many more) center around this typological division in the classification of writing systems. The basic problem of this division is that the kind of writing which the creators of universal languages seek will represent concepts and things without any phonological content, instead using only images to do so. Were such a system to be unearthed from the distant past, it would be classified as a proto-writing system, rather than as a writing system, because it has no phonological content. Even if such a semantic writing system were to have a sophisticated visual grammar, such a grammar would be very abstract, and could hardly be said to represent human speech. Several visual languages exist in the world which have sophisticated grammars, but none of them are considered writing. The two design problems, concerning the inherent ambiguity of icons and the proliferation of signs, were side-stepped by the ancients (unconsciously) through the use of what we call the rebus principle. Thus, an examination of the rebus principle and its mechanism may be useful to provide a solution to these design problems.
Most explanations of the rebus principle have examined the phenomenon mechanistically. There has been little theoretical reflection on what the use of visual puns might have meant to some of the first scribes of the world. The basic theoretical argument I advance in this paper is that we must understand the rebus principle as a scribal practice emerging from the phonosemantic intuitions of the first writers; That is, that it is best interpreted as a form of sound symbolism, and we can infer.

Phonosemantic theories of language do not have a high status in contemporary linguistics. This is due to the methodological issues in investigating any inherent link between the phonological composition of a lexeme and its semantic content. It is impossible to separate a phonosemantic link between sound and meaning in speaker intuition in any systematic way. De Saussure’s primary assumption about the signifier and the signified is that the link between the two is arbitrary. Any postulated link between the two is too subjective to lead to a scientifically valid theory. And yet many of the most famous linguists of the 20th century harbored phonosemantic intuitions.

The scientific status of phonosemantic theories of language is outside the scope of this paper. However, recently, a number of researchers from disciplines outside of linguistics, notably Margaret Magnus in computer science have attempted to develop rigorous methodologies for the study of phonosemantic phenomena in language, methodologies which are not too dissimilar from accepted analyses from phonology (Magnus 2001). In addition, recently Altmann (2023) has done work on obstruents and perceived sound-symbolic harshness in branding, while Pericliev (2022) has looked at lexical iconicity.

If one considers that the first scribes may have used rebii because they saw the words as being linked not just by an abstract pun, but by an essential link between sound and meaning (which in the writing
system becomes sound and image) we have a theoretical lens which can help us understand not just the historical process of the rebus principle, but also some of the special forms of writing which do not exist in alphabetic writing systems. In particular, phonosemantic assumptions may be the key to understanding why scribes in Egypt created inscriptions known in the Egyptological literature as “cryptographic” or “enigmatic”—inscriptions which creatively deploy images to represent phonological content in a mysterious way that is hard to interpret if one is an outsider (Cryptographic Writing citation). The art of cryptographic writing then becomes the manipulation of these two levels of representation to convey implicit messages by commenting slyly on the phonosemantic link between word and image, or the word and what it represents.

It can suffice to remain agnostic about the status of phonosemantics as a theoretical framework to explain language in its own right. What is useful for the purpose of inventing a universal semantic writing system is that it solves the two design problems from x’s paper by using the phonological level to clarify the ambiguities of the image layer in the writing system. From a design perspective, the use of a phonosemantic link between an image and a sound can serve as a mnemonic which helps the reader remember the phonological value of a character and mutually reinforces its pictographic meaning as well. It also allows the system to use existing language to express abstract concepts more in line with the original ideographic writing systems.

The use of a universal rebus principle may not at first be obvious. The problem with the use of the rebus principle is that it, by definition, obscures the meaning of an image in a code accessible only to those who speak the language it is written in. Thus a universal rebus principle would also seem to imply a lingua franca or the necessity of an Esperanto. However, as will be demonstrated, this may not be
necessary. The use of a “universal rebus principle” could be founded on a small lexicon of icons with a limited phonological content (such as, the characters of the IPA\(^1\)) which themselves represented semantic universals. Then the phonological content of the semantic universals, and the characters in the IPA, could each serve to clarify each other in the mnemonic sense implied by a phonosemantic interpretation of cryptographic inscriptions in ancient logographic writing systems. When this small set of universal concepts was combined to form other concepts, the results would be a lexicon which users of such a language could generate themselves, and to which they could provide their own meanings. The creation of such a system and its use will be outlined below.

## 2. Historical Critique

The idea of a universal language, historically, emerges in the context of the scientific revolution of the 17th century. Prior to that, schemes for languages tended to be either a search for the Adamic language—which typically centered around nationalist arguments for which tongue was spoken in the Garden of Eden—or the construction of Platonist or Angelic languages for the revelation of mystical or philosophical realities. It is the latter kind of language which forms the antecedent to the search for the universal languages of the Enlightenment. There is a direct line of descent of Wilkin’s *Real Character* from the thought of Bacon which was equally concerned with *hieroglyphs* as it was with science. Bacon calls hieroglyphs “real characters” by which he means characters which represent ideas by

---

\(^1\) The following abbreviations are used in this paper: IPA (International Phonetic Alphabet).
means of images. This idea would influence Comenius in his creation of the Orbis Pictus, a pictorial encyclopedia which still served to educate children well into the 18th and 19th centuries, and from Comenius had an influence on the early Royal Society, Wilkins, and through Wilkins, Leibniz and his characteristic universalis (Rossi 2013).

Bacon drew his ideas about hieroglyphs from a lively and colorful Renaissance tradition which saw them as non-discursive Platonist emblems, conveying information in an almost mystical fashion. These ideas were partly derived from the impressions of Neoplatonists who had written about hieroglyphs in classical antiquity. But they were largely influenced by the reception of a text written by a purported Egyptian priest, Horapollo Niliacus, who produced a dictionary of the meanings of Egyptian hieroglyphs (Niliacus 1993). Horapollo was an Egyptian priest who taught at the Serapaion of Alexandria, a temple to a syncretic Greek Egyptian deity, so he was teaching in a Hellenized context (Wildish 2017). What makes this text so provocative is that he provides the correct translations for many of the hieroglyphs in the first book. However, he does not give the actual linguistic reasons for the meanings of these glyphs. Instead, he gives allegorical explanations of the glyphs similar to a medieval bestiary, which reinforce Greek ideas about the hieroglyphs. But he was an Egyptian who clearly did have some genuine knowledge, and was working in the context of Neoplatonism. According to Wildish (2017), his student Philip translated this manuscript from Coptic into Greek and added a whole second book, where he invented a series of fake Egyptian hieroglyphs which never existed. Thus you have an ancient encyclopedia which is a complex mix of truth and fiction, partly Egyptian and partly a Hellenized vision of Egyptian culture (Iversen 1961). There is a chance that these allegorical ideas about hieroglyphs, at least in the
first book, actually did represent genuine Egyptian convictions about the meanings of their glyphs. The status of this interpretation of the glyphs depends on the Egyptians’ own attitudes towards writing, and to what degree the phonological and semantic layers were linked by the rebus, which is explicitly played with as in the case of so-called “cryptographic writing”. This interplay between sound and meaning, which is common to all ideographic writing systems, is of the utmost importance to interpreting how the reception of Horapollo influenced the debate around universal languages, and how we can subvert these mistakes in our own systems.

Horapollo’s encyclopedia was discovered by Cristoforo Buondelmonti, a Florentine explorer, in the 15th century on the island of Naxos. It was brought to Florence and triggered an enormous explosion of art in the European Renaissance (Iversen ibid.). To the intellectual of the early Renaissance, the belief was that the book taught the reader how to interpret Egyptian hieroglyphs and that all hieroglyphs were inscriptions in an allegorical mode. They extended this idea to Chinese writing, which the Jesuits were just beginning to encounter in those centuries. So these ideas about writing were extended to Chinese language, which was seen as similar, and many of these misconceptions have sadly continued among the general public.

This encyclopedia combined with already existing ideas of non-discursive writing in the Platonist tradition to create an idea of hieroglyphs as allegorical emblems, which generated a significant number of Renaissance “neo-hieroglyphic” inscriptions, including the Hypnerotomachia Polyphili as its finest exemplar (Colonna & Godwin 1999). These neo-hieroglyphs were supposed creations of a non-discursive writing system that was instantly understandable to the trained philosopher.
It is from this historical context that the Enlightenment inherits its idea of a perfect and universal language. There is simply a slight shift in emphasis. Rather than the creation of an ideal Platonist language which would reflect ideal forms or the speech of Adam, the universal languages of the Enlightenment sought to empirically derive a notation system which would have mathematical, logical and linguistic precision to represent the entities found by the new science (Kaye 1996). The spirit of the pursuit is similar, but its motivation in the Enlightenment is more earthly than it is heaven-oriented in the Renaissance. Yet the idea of real characters as representing unambiguously what they depict remains common throughout both, and this idea of reducing ambiguity through a semantic writing system comes down to us in our own day.

If this history—which is really the history of a Platonist reception of Horapollo and the idea of Egyptian writing in general—is to be more than just an excursion into a neglected episode in the history of writing, one must consider what relation the fundamental linguistic assumptions here have to actual ideographic writing systems, in light of the framework outlined in section one.

Were such a symbolic system as what Horapollo described realized it would be considered proto-writing by linguists today, and not true writing. Similarly, the Renaissance reception of Horapollo in the context of emblematic literature is also more akin to seal-stones and other proto-writing inscriptions from the ancient Near East than cuneiform and hieroglyphs. This conviction about the non-phonological nature of ideograms remains unchanged in the scientific revolution, although Bacon and Wilkins there is an obsession with reducing ambiguity which does not exist in the earlier hieroglyphic literature. Even where such linguistic schemes become conflated with mathematics and logic, as in Leibniz (1880), the pictographic or
symbolic emphasis remains. Yet pictograms and ideograms remain fundamentally ambiguous, and therefore such approaches can only yield syntactic logical operators similar to Frege’s begriffschrift and never the universal semantic language its creators aspire to.

In short, attempts to create universal semantic writing systems have been hampered by misunderstandings about the nature of ancient logographic writing systems, which, while they have led to creative misinterpretations that are highly productive, have ultimately made the creation of such systems impossible. What remains to be seen is whether an artificial rebus principle could be created which uses ideas from phonosemantics to clarify the ambiguities in pictograms.

3. Research Creation for a Universal Rebus Principle

Two approaches were attempted in assigning sounds to pictograms. The first looked at the frequency of character use across a Swadesh list of linguistic universals translated into roughly 2,000 unique natural languages. The second was to manually map the IPA onto a semantic taxonomy of those same universals according to a similar taxonomy for phonological features. The first approach was discarded, because there were fundamental methodological issues in how graphemes were mapped to specific phonemes in the dataset, which was derived from the PanLex project. However, the method did produce a new dataset of characters mapped to conceptual clusters and yielded some notable insights—such as the cross-linguistic use of the character ‘n’ for the concept ‘nose’. The frequency approach and the issues with it will be discussed in the first section. In the second section, the taxonomic approach and a resulting transcription will be presented.
3.1. Frequency Approach

The assumption of the phonosemantic rebus principle is that the number of concepts for a universal pictographic language need not be very large, but they should have as wide a linguistic coverage as possible. For this reason, interlingual dictionaries based on the largest possible linguistic dataset were desirable. The PanLex project was selected due to the enormous number of translations and language varieties contained in the dataset. The goal of the PanLex project was to build a lexicon of “every word in every language” and amounted to an attempt to digitize and ingest every translation dictionary in the world. Funded by the LongNow Foundation, the lexicon was used to create “Rosetta Disks”, micro-etched nickel disks with translations of the Universal Declaration of Human Rights transcribed in as many languages as possible, designed to last for 40,000 years, in order to create a language archive which would stand the test of time.

One of the resources produced for these artifacts was two Swadesh lists that were translated into as many natural languages as possible. One list contained 110 expressions translatable into 2,000 languages, and another contained 207 expressions translatable into 800 languages. To construct the pictographic writing system, expressions from both lists were used. The Swadesh lists were based on the work of Alexei et al. (2010) who extended Swadesh’s set of universal terms using Latin as a standard language. This was due to a perception by those researchers of Latin as being politically, culturally and ideologically neutral due to its status as a liturgical language and previous lingua franca of the West; In practice, Latin remains biased towards a Roman Catholic worldview and the perspective of Western culture, and it hardly politically or culturally neutral to serve as a base standard to be extended into 2,000 languages. The base list was then extended by PanLex to include as many translatable and verifiable words in as
The result is a dictionary of highly translatable concepts which were mostly common nouns, although there were some adjectives and verbs. The English translations of these concept clusters were input into Emojipedia to find a glyph that would represent them. While this was sufficient to find glyphs for most of the concepts in the dictionary, for some concepts it was necessary to use Gardiner’s sign list from Egyptian Hieroglyphs to find appropriate Unicode representations of the concept clusters. Thus due to the fact that the concepts contained in the dictionary also tended to be very simple—such as “egg”, “fire” or “water” it was relatively simple to find pictographic representations for the less abstract concepts.

Next, a simple script was written which ranked the characters used to write each concept by frequency. This produced an irregular dataset with differing rows per column, as different distributions of characters are present for each concept. The result was a ranked list of letters for each concept in Swadesh 110. For most concepts, the most common letter used to write them was “a”, “u” or “i”, with some exceptions. As mentioned above, the use of the nasal consonant ‘n’ for the concept of nose was significant.

However, there are significant issues in interpreting the ground truth of the data, even before statistical methods would be employed to search for potential patterns. Due to the proliferation of Western scripts, the Roman alphabet was vastly more overrepresented than any other script. While this aided in English speakers interpreting the data, significant phonological implications that may have been kept in other, less widely represented scripts, were at risk of being lost. Furthermore, it was difficult to tell which characters represented the Roman versions of a glyph and which were Unicode characters for identical glyphs from another writing system, such as the IPA. This
led to a further ambiguity of interpretation that would necessitate examining the underlying Unicode encoding to disambiguate.

These issues in the data are related to how PanLex ingested wordlists (dev.panlex.org) (Kamholz et al. 2014). Due to the fact that PanLex was attempting to reach the widest possible coverage in the world, any and all translation resources were used. Many word wordlists from fieldwork done many decades previously, where fieldworkers did not always have time to make a rigorous IPA transcription of a text, and used idiosyncratic transcription conventions often relying on the Roman alphabet.

Even were this significant, the data represent only graphemes, and not phonemes, and making inferences about phonological relations to conceptual clusters without rigorous IPA transcriptions for all the data is questionable at best.

This makes the data very difficult to understand, nevertheless, it is attached here as an important part of the process of assessing the relation of sound to meaning cross-linguistically and may be of interest to other researchers.

### 3.2. Taxonomic Approach

A taxonomy of concepts was created from those that were more abstract, down to those that were more concrete and concerned with the human world. Weak sub-class relations bound mother concepts to daughter concepts, with an interpretation of sub- and super-class relations as being similar to metaphor or analogy. Each glyph was assigned an antonym or opposite, which was based on the structure of the original Swadesh list which had pairs of opposites. Each glyph was also given a sister, or a complement, from a concept that was judged to be similar to it but marked in a way that did not indicate opposition. This taxonomy formed a binary branching tree pictured in
Then, in order to create a phonosemantic rebus for each interlingual glyph, characters from the international phonetic alphabet were mapped to each glyph. Characters that were marked for similarity were chosen as complements, and where possible phonological features which were opposite each other in the vowel trapezium for vowels or the sonority hierarchy for consonants were selected to represent opposition. The mapping of abstract concepts—natural features—common human objects from the conceptual taxonomy was mapped to vowels—semivowels—consonants in the phonological taxonomy. This was also inspired by the work of Agrawal (2020).

This led to a set of three taxonomies when visualized, one which had terms written in English in natural language (Figure 1), one which had emoji representations of those terms (Figure 2), and one which had a mapping to IPA characters (Figure 3). This functionally created an equivalence between IPA characters and emoji. The final implementation was to transcribe an English text in Emoji using the IPA mappings. For this exercise, the Universal Declaration of Human Rights was translated into IPA using an automated IPA translator. A script was then written which mapped the text to emoji using the mappings from the taxonomy. This produced a text where a phonetic transcription of English was mapped to emojis. Figure 4 shows this new “Rosetta” transcription in the Roman Alphabet, IPA, and the Emoji mapping.
The set of primitives from the Swadesh list was arranged into a taxonomy representing their semantic relations, with the most abstract terms placed in the center. Terms that were antonyms were placed across the circle from each other. Related terms were placed as children or sisters of other terms in a binary branching tree.
Finally, a combination of emojis and hieroglyphs was used to create iconic depictions of the concepts in the Swadesh list.
According to phonetic features, the characters of the International Phonetic Alphabet (IPA) were arranged hierarchically. The innermost rings include vowels, followed by semivowels, and the outermost ring include consonants. This is to mirror the abstract/concrete divide of the concepts in the Swadesh list. Sounds which are related by phonological features, particularly +voice/–voice for consonants are grouped as sisters in the outermost ring: relations between vowels in opposition are based on the vowel trapezium.
Prototype version of the universal declaration of human rights mapped to emoji using a toy “universal” rebus principle. This rosetta inscription allows the transcription of phonetic data with semantically keyed characters.
4. Discussion

The goal of this creative experiment was to make a writing system composed of pictorial representations of universal concepts which could in principle be used to write phonetically in any language. The idea of a “phonosemantic rebus principle” as applied here is mostly a metaphor. There is no attempt made to justify that these sounds actually do correlate with their meanings, and the work has instead been a creative act of making associations. Perhaps in the future, it would be possible to find an alphabet of phonosemantic primitives, but scientifically this will probably be impossible. Regardless, as a productive practice for producing a writing system, taking a “phonosemantic rebus principle” as a metaphor has a number of distinct advantages.

While a simple writing system has been achieved the question still remains how readable such transcriptions are, and what use they might possibly have. A challenge in the transcription which was produced is it is difficult for English speakers trained in the Roman alphabet to recognize as English, as it contains a true phonetic transcription of the language in IPA. Even if this writing system were truly phonosemantically universal it remains not very accessible to readers. Simpler phonological mappings which are maybe less technical than those used by linguists could potentially create a writing system which is easy to remember and use.

Nevertheless, a phonosemantic approach to making a pictographic writing system might seem strange in light of the goals of Enlightenment interlingua. Such a writing system surely would only obscure differences between cultures, as text written in each language would only be readable by those who already spoke that language. I argue that there are two distinct advantages in this approach: mnemonic, and
productive.

The mnemonic advantage of the phonosemantic approach is that the pictorial glyph ought to help the reader remember the phonological meaning, and the phonic meaning should in turn reinforce the semantic in the reader’s mind. Regardless of the scientific status of phonosemantics as a method of linguistic inquiry, thinking in phonosemantic terms is common and intuitive for speakers across all languages. Making an association between sounds and what things look like will help readers keep the meanings of the glyphs in memory—especially if future research can find a more rigorous way to determine how sounds are shared in conceptual categories across languages.

The productive aspect bridges this project to create a scientific taxonomy and writing system with projects for making universal spoken languages like Esperanto, although it implies no set vocabulary or grammar as yet. The use of pictograms to create strings of concepts connected logically would, with this system, also imply the creation of phonetic words. These words, when combined with the mnemonic aspect of the system, could lead to an endlessly productive linguistic game allowing the writer to invent their own words and puns across languages, which could be endlessly generative and interpretable to someone familiar with the system. While no grammar for such a language would be proposed, similar to the process of creolization, the idea would be to leave the emergence of a grammar to the consensus process which would be negotiated as people use and learn this writing system.

While such a writing system would not be instantly universal by a top-down, hierarchically organized fiat, through the use of centuries perhaps its speakers and writers could gradually create a global creole which would be keyed to universal concepts which could be used,
where necessary, in the mathematical and logical formations we find in the sciences. This work and these brief sketches are meant to open the conversation about an alternative way of creating universal languages which is more historically grounded in how ideographic scripts actually functioned.

References

3145–3150. Reykjavik: Harpa Concert Hall.


