A Computational Model of the Discovery of Writing[1]

Richard Sproat
Google, Inc.

Introduction

The genesis of writing starting about 5,200 years ago, much like the genesis of language tens or hundreds of thousands of years before it, is the stuff of conjecture. While there are many ideas on how humans first started using vocal sounds to communicate complex thoughts, and how eventually they learned to transfer those sounds to written media, there is a dearth of hard facts. In the case of the origin of writing, it has become clear that the token theory of the development of writing in Mesopotamia, most notably promoted by Schmandt-Besserat (Oppenheim, 1959; Schmandt-Besserat, 1996), can only be a part of the story (Woods et al. 2010, pp. 48-49), and we have little if any hard evidence for the remaining pieces.

Compounding the problem is that while there have been hundreds of writing systems developed over the past 5,200 years, some within living memory, only in four parts of the world — Mesopotamia, Egypt, China and Mesoamerica — did writing apparently develop *ex nihilo*. Writing is a product of civilization, but it is not a *necessary* product of civilization: many civilizations, from the Incas of South America, to the Aztecs of Central America, to the Indus Valley to the Gojoseon Dynasty of Korea, have lacked anything clearly identifiable as writing, though they certainly had other notational systems. Were the four cultures that did develop writing smarter than the others? Or (more likely) did certain properties of their culture or their language make it more likely that they would develop writing? With only four instances to work with, it is hard to make any robust claims.

One way around this limitation is computational simulation. Simulations have already been used extensively in the modeling of historical change in language, such as the spread of linguistic features in social networks, and the emergence of linguistic properties (e.g. Kirby, 1999; Niyogi, 2006; Steels, 2012).

A complete simulation of the evolution of writing would aim to model the following factors, among others. First of all, the types of nonlinguistic symbol systems in use in the culture, and the existence of combinatorial systems where symbols occur in "texts". Most if not all cultures use symbols to represent concepts that are, because they do not depend on language, *nonlinguistic,* and many of these are complex systems where whole texts can be "written" in these symbols, so that the system resembles true writing in many ways (Sproat, 2014); see Figure 1 for some examples. Some of these, such as the accounting system in early Mesopotamia did eventually

[1] Presented at the *Signs of Writing* conference, Oriental Institute, University of Chicago, November 2014.

develop into writing, but most did not, and it is a reasonable guess that some of them *could not have* developed into writing.

[INSERT FIGURE 1 ABOUT HERE]

A second class of factors is economics or other social properties that would encourage the development of better means of record keeping. Wang (2014) discusses the development of record keeping and its role in the recording the myths of the state as well as the accounting needed to keep the state functioning, and he compares societies — Mesopotamia, China and Mesoamerica — which developed and used writing for these purposes, with societies — most notably the Inca — which did not.

Thirdly, linguistic properties have been argued to be relevant (Daniels, 1992; Boltz, 2000; Buckley, 2008). The key insight needed for full writing is the realization that a symbol that had been used to represent an idea, could also be used to represent the sound of a word or morpheme associated with that idea. For this to work, it helps if the language used by the would-be scribes is one where it is relatively easy to find homophones or close homophones. Puns, in other words, should be relatively easy to make. Alternatively the notion of what it means for something to *sound like* something else might be more relaxed in some languages than others: for example if a language has a Semitic-style root-and-pattern morphology where vowels change drastically in related words, a word such as *patak* might count as sounding like *pituk*.

Finally, for writing to become practical and widespread, there must be relatively lightweight materials available as writing surfaces (Farmer et al. 2002), with lightweight devices such as styluses, brushes or pens to incise or inscribe on the surfaces. A standing joke in the *Astérix* comics has the Romans writing on slabs of marble with a hammer and chisel: obviously such a system would not be practical for everyday use.

Simulating the relevance of the type of non-linguistic system, the social and economic factors, the linguistic factors and the physical properties of the writing surfaces is clearly a tall order. But we can make some progress on some parts of the problem.

In this paper I concentrate on the linguistic factors and report on a preliminary system that simulates the realization, discussed above, that a symbol that had heretofore been used to represent an idea, could also be used to represent the sound of a morpheme associated with that idea.

The simulation starts with symbols randomly assigned to 100 basic concepts, as well as up to 1,000 morphemes whose meanings can be decomposed into one or more of these basic concepts.

For any concept there may be several morphemes that can be said to denote that concept: for example *human*, *person* and (one sense of) *soul* denote approximately the same concept in English. In the simulation, one of these morphemes is randomly picked to become the one associated most strongly with the concept's symbol. Now that the symbol is logographic, and thus lexically and not just conceptually grounded, the transfer of the symbol to other morphemes with a similar sound is licensed. At the same time, the semantic basis of the symbols is preserved so that morphemes that have related semantics may also use symbols on the basis of their meaning. The result is a system that writes morphemes either using purely semantic elements, a mixture of semantic and phonetic elements, or purely phonetic elements — in other words a system that resembles all ancient writing systems.

We turn now to a description of the model.

The model

The model is implemented in Python and consists of about 500 lines of code. It uses the *pyfst* interface (http://pyfst.github.io/) to OpenFst (Allauzen et al, 2007) and depends on various grammars written in the *Thrax* open-source finite-state grammar compiler (Tai et al, 2011; Roark et al, 2012).

Phonotactics

The model allows for the setting of a number of different parameters. The first is the basic pattern of morphemes in the language, with the possible setting being monosyllabic or (maximally) disyllabic. We define a syllable as follows:

σ = C? V R? C?

where C and V are consonants and vowels, respectively, and R is a sonorant. These in turn are defined as follows;

C = p, t, k, b, d, g

V = a, e, i, o, u

R = N, r

where N is an unspecified nasal. The phonotactic grammar is written using Thrax.

For a language with monosyllabic morphemes, we randomly generate about 1,000 monosyllables from the grammar. For a language with maximally disyllabic morphemes about 1,000 mono- or disyllabic forms are generated.

Phonetic Similarity

The second parameter is the amount of phonetic similarity between two morphemes *m* and *n* such that a symbol already used to represent *m* can be adopted for *n.* Put in terms of a more widely used term, this setting parameterizes the specific way in which the *rebus principle* works in the developing system. This is also defined by means of a grammar and currently has four possible settings:

1. CLOSE:
    - obstruents match on place of articulation
    - vowels must match
    - sonorants optional
    - Example: *daNg* matches *tak*
2. CLOSE-RHYME:
    - as above, but "rhyming" is sufficient
    - Example: *daNg* matches *bak*
3. CLOSE-V-FREE:
    - same as CLOSE, but vowels don't need to match, and V matches Ø
    - Example: *donek* matches *dink*
4. STRICT:
    - Exact match required

Association of Concepts with Morphemes

The system starts with 100 basic concepts, shown in Figure 2. Many of are these easily depictable, and are represented by glyphs in various ancient writing systems.

[INSERT FIGURE 2 ABOUT HERE]

In general a concept may be linguistically expressed by more than one word. For example, the concept of *human being* can be expressed in English by several words, including *human, person,* and one sense of *soul.* Therefore in the simulation between one and three morphemes are associated with a given concept, with one of these being selected as the primary way of expressing the concept: in English, *person* is probably the most natural way to express the concept *human being* in everyday language.

This assignment uses up a portion of the whole set of generated morphemes. The remaining morphemes are associated with combinations of two or three concepts. Some examples of such

random combinations can be found below:

CAVE, TEMPLE, GOAT

GRASS, DEMON

WIND, FINGER

MOON, LEAF

While some of these combinations may seem silly, note that seemingly odd combinations of concepts are manifest, for example, in many Sinitic characters. For example, it is not *a priori* obvious what insects have to do with rainbows, but the *insect* radical 虫 is used to write the character 虹 *hóng* 'rainbow'. Similarly while it is clear why the Japanese *kokuji* (native Japanese character) 嬶 *kakā* '(vulgar word for) wife' contains the female radical 女, it is not obvious why the other component is 鼻 'nose'.

Symbols

Each basic concept is randomly associated with a symbol. For these we use dingbats and other mostly non-script symbols from the Unicode Basic Multilingual Plane. Some examples of symbols used can be found in Figure 3. Note that in the work reported here, no attempt is made to make the symbol graphically reasonable for the concept being depicted.[2]

[INSERT FIGURE 3 ABOUT HERE]

Running the model

In order to run the model, one must choose settings of the parameters described above. In particular we choose combinations of monosyllabic versus disyllabic morphemes and phonetic closeness, or in other words we choose one of the phonotactic settings under 1 below, and pair it with one of the phonetic match settings from 2. A complete set of runs would include all combinations of these.

1. **Phonotactics:** MONOSYLLABIC / DISYLLABIC
2. **Phonetic Match:** CLOSE / CLOSE-RHYME / CLOSE-V-FREE / STRICT

[2] In ongoing work on a more recent simulation than the one reported here, we try more carefully to choose symbols that make some graphical sense.

For each combination we randomly generate five "languages" -- i.e. a set of roughly 1,000 morphemes dictated by the phonotactics, and then run the simulation for each of these languages five times, keeping the phonetic match setting fixed.

The steps in the simulation, some of which have already been described above, can be given informally as follows:[3]

1. Assign simplex concepts and symbols to morphemes.
2. Randomly assign complex concepts to remaining morphemes.
3. Select the *primary morpheme* for a concept as the one that inherits that concept's symbol.
4. Assign spellings to morphemes that do not have one:
    a. Symbols associated with shared semantic components, and/or
    b. Symbols associated with similar sound according to the predefined notion of phonetic similarity — the rebus principle.
5. Extend the phonetic coverage with telescoping:
    a. ba + ad → bad
6. The result is a mix of semantic and phonetic components.

The first crucial step in the symbols system becoming linguistic is step 3, where a symbol becomes associated with a particular morpheme, rather than with a concept. At that point the system can be characterized as *logographic*. But the key insight, what one might term the "eureka moment", comes at step 4b, when it is realized that symbols may be used in *rebus* fashion purely for their sound values.

The connection between the meanings of morphemes and their sounds is of course already present as a critical part of spoken language understanding and production. So the phonological recoding of previously logographic symbols is somewhat natural, since it depends in part on this connection. That such recoding is natural — at least with fluent readers of a writing system — can be seen in the fact that it modern writing systems that have a significant logographic component, one nevertheless finds evidence for a direct mapping path between the symbol and its pronunciation that effectively bypasses the semantics. Among modern writing systems, Japanese *kanji* surely hold the best claim to be truly logographic, particularly in their use to represent native morphemes. Figure 4 gives some examples of a set of characters involving the phonetic component 里. In Mandarin, all of these characters are in fact homophones meaning that the phonetic component in this instance is a very useful guide to the pronunciation. However, in their use to represent native Japanese morphemes, the phonetic component is

[3] For those interested, the code (which is, however, evolving) can be made available upon request.

useless, since the morphemes all have very different pronunciations. Therefore the system is effectively logographic: the symbol represents a morpheme, with no indication of the pronunciation.

[INSERT FIGURE 4 ABOUT HERE]

Nevertheless, Japanese readers treat *kanji* as if they represented sound. One piece of evidence for this is from writing errors. For example, Horodeck (1987), in a study of Japanese spelling errors, found that many errors involve substituting a character with the wrong meaning, but the right sound. While most of his examples involved Sino-Japanese (*on*) readings, where the phonetic component of the character often provides a clue to the pronunciation as in Chinese, 7% involved native (*kun*) readings.

Matsunaga (1994) found similar results in an eye tracking experiment where she measured participants' fixation rates on intentional spelling errors in text. The more noticeable an error is, the longer a subject will fixate on it. Significantly, subjects had much lower rates of fixation on errors that involved substitutions of homophonic characters, than they did on errors that involved substitutions of non-homophonic characters. In other words, when an error involves a character that sounds the same as the correct character, it is much less likely to be detected. This only makes sense if fluent readers Japanese have "phonologized" characters: even though from the point of view of the typology of writing systems, *kanji* are clearly logographic, what matters for skilled reading is a rapid mapping between a glyph and its sound.

In summary, even if a written symbol formally represents a word or morpheme, it is virtually automatic among users of the writing system to reinterpret that symbol as representing the sound of the word or morpheme. While the situation among preliterate "readers" in ancient civilizations was surely different — for one thing they would presumably not have been as fluent at reading as present-day readers — it nonetheless seems likely that the same cognitive connections would have been at play, making it natural to use a symbol for its sound in addition to its original meaning.

Results of the simulation

The result of running the algorithm is a set of morphemes associated with semantic features, and, in most cases, with a spelling consisting of a set of one or more glyphs. We start off by giving some examples of some of these from a couple of the different parameter settings.

[INSERT FIGURES 5 AND 6 ABOUT HERE]

Consider first of all Figure 5, which presents some examples of morphemes and their spellings in one of the MONOSYLLABIC, CLOSE runs. Glyphs enclosed in "S{}" represent semantic components, and

those in P{} represent phonological components. In the sample shown, some morphemes are represented only by semantic markers, whereas in others they are represented by semantic-phonetic compounds similar to Chinese *xingsheng* characters. Figure 6 shows some of the phonetic components that emerged and their various phonetic values.

[INSERT FIGURES 7 AND 8 ABOUT HERE]

Figure 7 shows a similar set of examples from one of the DISYLLABIC, CLOSE-V-FREE conditions. In these cases some of the morphemes are represented purely phonetically (only having a P{} element not an S{} element). Figure 8 shows some of the phonetic elements. Note that many of the phonetic elements are used to represent morphemes that have different vowels. In the series for the sign <¯> /dipirg, taNpak, dubog, dirbik**/,** for example, one can characterize the spellings as representing /TVR?PVR?K/, where /T/ is an apical obstruent, V is any vowel (or in general no vowel), /R/ is a sonorant, /P/ is a labial obstruent and /K/ is a velar obstruent. Thus <¯> represents a disyllabic form where the vowel is unspecified, sonorants are optional, and only the place of articulation of the obstruents is fixed. It is interesting to consider this in light of Gelb's (1963) well-known view that Semitic consonantal writing was a syllabary with a wild-card vowel. Most later authors on writing systems (including myself; see Sproat, 2000) have argued that Gelb's analysis was largely an attempt to shoehorn consonantal systems into his teleological theory that had writing systems evolving from logographic systems, to syllabaries and eventually to the alphabet. While this is probably a fair characterization of his reasons for analyzing consonantal writing in this way, the mechanism by which the simulation arrives at the sets of phonetic values in Figure 8 suggests that after all Gelb may have been partly right.

[INSERT FIGURE 9 ABOUT HERE]

Figure 9 give some statistics on the overall properties of the systems developed under the different conditions. Not surprisingly the smallest proportion of morphemes that end up with a spelling occur in conditions that require strict phonetic matching (MONO, STRICT; or for consonants DISYLLABIC, STRICT-V-FREE). DISYLLABIC, CLOSE also results in a smaller number of spellings and also the smallest proportion of phonetic spellings of all the systems. Evidently it is much more difficult to develop a phonetic writing system for a disyllabic language if one insists on a close match, with exact match for vowels. If on the other hand vowels are allowed not to count, then nearly all morphemes in the disyllabic language end up being writeable and 50% of them have a phonetic spelling.

The MONOSYLLABIC-CLOSE condition, which is arguably nearest among all our simulations to the situation in Ancient Chinese, has the interesting property that about 30% of the spellings of morphemes are semantic-phonetic spellings. This is close to the rate of 34% reported by

DeFrancis (1984, Table 3, p. 84) for semantic-phonetic characters in Shang Dynasty Oracle Bone texts.

In summary, the simulations show that, as has been suggested by previous scholars (e.g. Daniels, 1992; Boltz, 2000; Buckley, 2008), there is indeed a dependence on the phonological structure of the language but also, not surprisingly, on how phonetically "close" one morpheme must be to "sound like" another. Monosyllabic languages seem to have an advantage over languages that have longer morphemes. But loosening the notion of phonetic closeness to allow vowels not to count, as would be motivated in a language with an Afroasiatic-style root-and-pattern morphology yields greater success in languages with disyllabic morphemes. Also clear is that requiring *strict* matching is not a winning strategy, and in fact strict homophony never seems to have been a requirement for extending the use of symbols in early writing systems.

Synopsis and future work

It is worth remembering that while we may well have evolved to speak, we did not evolve to read. As DeHaene (2009) argues, the occipitotemporal region of the brain in the left visual cortex, part of the apparatus for identifying objects (DeHaene, p. 125), has been co-opted in literate people and used for the low-level processing of scripts. It does not seem to matter what script is involved: it could be the Roman alphabet, the Hebrew alphabet, or Chinese characters. In any case the initial processing of writing passes through this co-opted region that DeHaene terms the brain's *letter box.* From there information flows to the parts of the brain that deal with sound and ultimately with meaning, parts that are also active in illiterate speakers.

Somewhere around 5,200 years ago the identical co-opting of the same region of the brain must have taken place when the very first writing systems were developed in Mesopotamia and Egypt. The connections between the newly created *letterbox* and the parts of the brain that were already active in speech were certainly not immediate. In particular, for the phonetic connections, as noted in Woods et al. (2010, p. 20) it was not for a few hundred years, in the first quarter of the third millenium, that "rebus writings would play a significant role" — and "not until the second half of the third millennium that the linear order of signs reflected sequential speech." Similarly in Egypt, where phonetic signs were more important early on, nonetheless "over five hundred years would pass before the script recorded continuous speech." Still, "[t]he rebus principle was obviously known at an early date in both [the Mesopotamian and the Egyptian] systems, and so the potential to represent speech was there from the very beginning, or nearly so." Thus the mechanism assumed in the simulations reported here, namely the connection between the sound of a morpheme and a symbol representing it, was something that was available from early on in the development of writing. If its widespread application was not as immediate as our simulations suggest it should have been, it is likely that this is because in the earliest phases of writing there was no perceived need to extend the system beyond its logographic core. In our simulations there is an assumption that one *wants* to try to write as many morphemes as possible, but of course this desire to write ever more morphemes should really be a parameter of the system: if all one uses a graphical system

is for recording amounts of a small set of commodities, for example, there would be little motivation to extend it.

In any event, some of the connections between the *letterbox* and the remainder of the linguistic system were likely to have been helped or hindered by properties of the languages involved, as suggested by the simulations reported here.

In future work, we will further develop these models to explore this and related questions in greater depth, and we will provide further quantitative evidence for how well the system models true early writing systems.

References

Cyril Allauzen, Michael Riley, Johan Schalkwyk, Wojciech Skut and Mehryar Mohri, "OpenFst: A general and efficient weighted finite-state transducer library," *Implementation and Application of Automata,* Lecture Notes in Computer Science, Volume 4783. (Berlin: Springer Verlag, 2007), pp. 11-23.

William Boltz, "Monosyllabicity and the origin of the Chinese script," Preprint 143. (Berlin: Max-Planck-Institut für Wissenschaftsgeschichte, 2000).

Eugene, Buckley, "Monosyllabicity and the origins of syllabaries," Paper presented at the annual meeting of the Linguistic Society of America, 2008. (http://www.ling.upenn.edu/~gene/papers/monosyllabicity.pdf.)

Mark Changizi and Shinsuke Shimojo, "Character complexity and redundancy in writing systems over human history," *Proceedings of the Royal Society B.* 272 (2005): 267–275.

Peter Daniels, "The syllabic origin of writing and the segmental origin of the alphabet," in *The Linguistics of Literacy* edited by Pamela Downing, Susan Lima and Michael Noonan, Typological Studies in Language 21 (Amsterdam and Philadelphia: John Benjamins, 1992), pp. 83-110.

John DeFrancis, *The Chinese Language: Fact and Fantasy*, (Honolulu: University of Hawaii Press, 1984).

Stanislas, DeHaene, 2009. *Reading in the Brain: The Science and Evolution of a Human Invention,* (New York: Viking, 2009).

Steve Farmer, John Henderson and Peter Robinson, "Commentary traditions and the evolution of premodern religious and philosophical systems: A cross-cultural model," 2002, revised version in press.

Ignace Gelb, *A Study of Writing,* 2nd Edition (Chicago: Chicago University Press, 1963).

Richard Horodeck, "The Role of Sound in Reading and Writing Kanji," (Ph.D. diss., Cornell University, 1987).

Simon Kirby, *Function, Selection and Innateness: the Emergence of Language Universals,* (Oxford: Oxford University Press, 1999).

Sachiko Matsunaga, "The Linguistic and Psycholinguistic Nature of Kanji: Do Kanji Represent and Trigger only Meanings?" (Ph.D. diss, University of Hawaii, 1994).

Partha Niyogi, *The Computational Nature of Language Learning and Evolution,* (Cambridge: MIT Press, 2006).

A. Leo Oppenheim, "On an operational device in Mesopotamian bureaucracy," *Journal of Near Eastern Studies,* 18 (1959): 121-8.

Brian Roark, Richard Sproat, Cyril Allauzen, Michael Riley and Jeffrey Sorensen, "The OpenGrm open-source finite-state grammar software libraries," *Association for Computational Linguistics (System Demonstrations)* (2012), pp. 61-66.

Denise Schmandt-Besserat. *How Writing Came About,* (Austin: University of Texas Press, 1996).

Richard Sproat, *A Computational Theory of Writing Systems*, ACL Studies in Natural Language Processing Series, (Cambridge: Cambridge University Press, 2000).

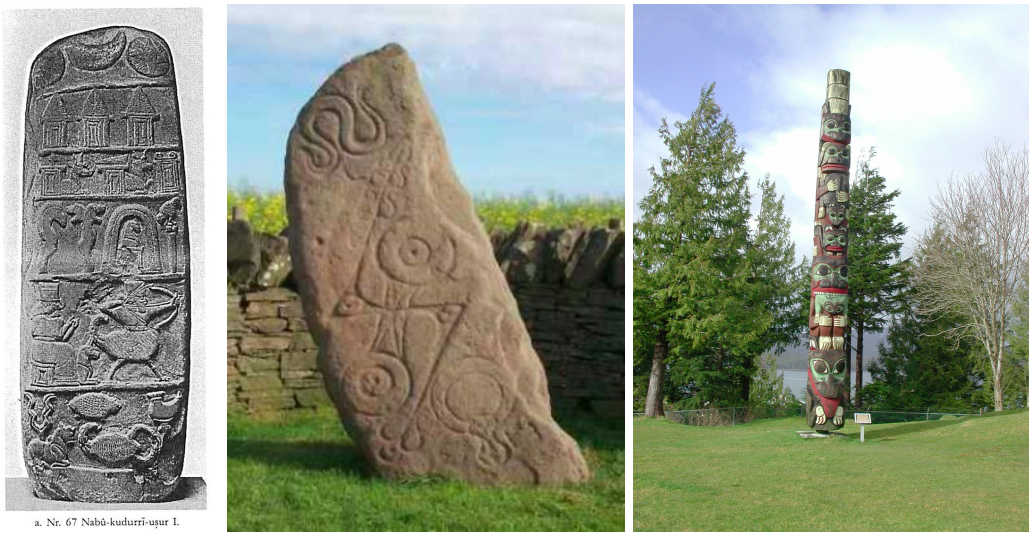
Luc Steels, Luc, *Experiments in Cultural Language Evolution*, (Amsterdam: John Benjamins, 2012).

Terry Tai, Richard Sproat and Wojciech Skut. 2011. "Thrax: An open source grammar compiler built on OpenFst." *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop* (2011).

Haicheng Wang, *Writing and the Ancient State: Early China in Comparative Perspective,* (Cambridge: Cambridge University Press, 2014).

Christopher Woods, Emily Teeter and Geoff Emberling (editors), 2010. *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond,* Oriental Institute Museum Publications, 32, (Chicago: Oriental Institute, 2010).

Figures



a **b** **c**

Figure 1. Some combinatorial non-linguistic symbol systems:

a. Babylonian *kudurru* stone, Kassite dynasty, ca 1125-1100 BC. From, Ursula Seidl, *Die babylonischen Kudurru- Reliefs. Symbole mesopotamischer Gottheiten,* (Freiburg: Universitätsverlag, 1989), Plate 23a. Used with permission.

b. Pictish symbols: Aberlemno stone, Scotland, 6th-8th century AD (http://www.stams.strath.ac.uk/research/pictish/database.php).

c. Totem poles, North American Pacific Northwest, 19th-20th century (From http://www.ourbc.com/, used with permission).

PERSON, MAN, WOMAN, HOUSE, BRONZE, SWORD, MEAT, SHEEP, OX, GOAT, FISH, TREE, BARLEY, WHEAT, WATER, STONE, CLAY, THREAD, CLOTHING, FIELD, TEMPLE, GOD, AXE, SCYTHE, DOG, LION, WOLF, DEMON, SNAKE, TURTLE, FRUIT, HILL, CAVE, TOWN, ENCLOSURE, FLOWER, RAIN, THUNDER, CLOUD, SUN, MOON, HEART, LUNG, LEG, ARM, FINGER, HEAD, TONGUE, EYE, EAR, NOSE, GUTS, PENIS, VAGINA, HAIR, SKIN, SHELL, BONE, BLOOD, LIVER, FARM, LOCUST, STICK, STAR, EARTH, ASS, DEATH, BIRTH, WOMB, MILK, COAL, SEED, LEAF, CHILD, ANTELOPE, BEAR, BEE, MOUSE, DUNG, PLOUGH, SPROUT, ICE, DAY, NIGHT, WINTER, SUMMER, AUTUMN, SPRING, KING, GOOSE, PRIEST, ROAD, CART, GRASS, FIRE, WIND, NAIL, BREAST, BOWL, CUP

Figure 2. 100 basic concepts.

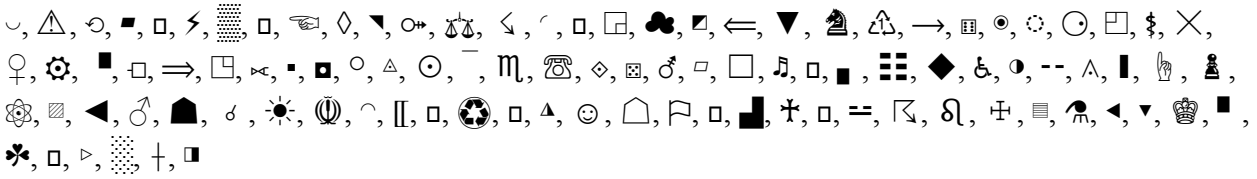


Figure 3. Examples of symbols used.

	Ma.	*Ja. (kun)*	*Meaning*
里	*lǐ*	*sato*	⅓ mile, village
鯉	*lǐ*	*koi*	carp
裡	*lǐ*	*ura, uchi*	inside
理	*lǐ*	*suzi, kotowari, osameru*	reason
厘	*lí*	*mise*	1000th of a tael

Figure 4. Some examples of Japanese characters with their native (*kun*) reading, and the corresponding pronunciation in Mandarin.

toN	PERSON	S{  }
kerg	PERSON	S{  }+P{☆}
gib	PERSON	S{  }+P{}
kak	MAN	S{  }
gab	WOMAN	S{  }
pet	WOMAN	S{}+P{}
urb	HOUSE	S{▲}
kuNt	HOUSE	S{▲}+P{}
a	BRONZE	S{☞}
dep	BRONZE	S{☞}+P{☞}
keg	BRONZE	S{☞}+P{☆}

Figure 5. Examples of some spellings arrived at in the MONOSYLLABIC, CLOSE condition

 tut

 bug

 dek

 eNp eNb ep eb

 boNb pob

 ked get

 girt kiNt gid

 beb

 god kod

 tot

 tok

Figure 6. Some phonetic elements derived in the MONOSYLLABIC, CLOSE condition. The right hand column shows the phonetic values that these elements represent. Those involving two glyphs were derived by telescoping (see above).

kape	PERSON	S{  }
dekik	PERSON	S{}+P{  }
akkiN	PERSON	P{}
buNbuNg	MAN	S{}
dadeNt	MAN	P{}
to	MAN	S{}+P{}
gea	WOMAN	S{}
tebe	HOUSE	S{}
eog	BRONZE	S{  }
geord	BRONZE	S{}+P{  }
appu	BRONZE	P{  }
akug	SWORD	S{  }

Figure 7. Some spellings arrived at in the DISYLLABIC, CLOSE-V-FREE condition.

uep oba eorp pai

oguk karig **ogek**

dipirg taNpak dubog dirbik

tugub

pebi obpe puuNp

torkga

okep keap kiNb kiub gaeNb

durap tepoN dube taup dapaN tuob

dibip

aduNg atog

Figure 8. Some phonetic elements in the DISYLLABIC, CLOSE-V-FREE condition. The right hand column shows the phonetic values that these elements represent. Note the cases in boldface showing vowel alternations.

	# entries	w/ spelling	w/out spelling	phon	sem+phon	purephon	sem
DI, CLOSE	661	462 (0.70)	198 (0.30)	47 (0.07)	35 (0.05)	11 (0.02)	451 (0.68)
DI, CLOSE-RHYME	648	508 (0.79)	139 (0.21)	127 (0.20)	100 (0.15)	26 (0.04)	482 (0.74)
DI, CLOSE-V-FREE	644	640 (0.99)	3 (0.01)	324 (0.50)	262 (0.41)	61 (0.10)	579 (0.90)
DI, STRICT-V-FREE	649	504 (0.78)	144 (0.22)	112 (0.17)	87 (0.14)	24 (0.04)	481 (0.74)
MONO, CLOSE	643	580 (0.90)	62 (0.10)	239 (0.37)	193 (0.30)	45 (0.07)	534 (0.83)
MONO, CLOSE-RHYME	651	630 (0.97)	20 (0.03)	310 (0.48)	253 (0.39)	56 (0.09)	574 (0.88)
MONO, STRICT	640	487 (0.76)	152 (0.24)	95 (0.15)	71 (0.11)	23 (0.04)	464 (0.73)

Figure 9. Results of the simulations averaged over 25 runs under each of the parameter settings. The left-hand column gives the parameter setting. The remaining columns give: the mean number of morpheme entries in the simulation; the mean number and proportion that end up with a spelling; the mean number and proportion without a spelling; the mean number and proportion of spelled morphemes that have a phonetic component; the mean number and proportion that have a semantic-phonetic spelling; the mean number and proportion that have a purely phonetic spelling; and the mean number and proportion that have a semantic spelling.