

Journal of Universal Language 16-1
March 2015, 87-130

'Real Character': Towards a Lexicon

Ian Maun

University of Exeter, United Kingdom

Abstract

This paper enlarges on previous work undertaken on a modern 'Real Character,' that is to say, a pasigraphical form of writing in which meanings are conveyed digitally to the reader by means of icons, indices and symbols (in the Peircean sense) without the intervention of phonetically-based writing. The purpose and audience for such a system are examined. The basis for a lexicon is sought in work on simplified language and possible semantic universals, and the process of affixation in natural languages is examined as a way of identifying prominent semantic notions. Estimates of necessary vocabulary size are made on the basis of previous studies. A critique of a number of existing visual languages is given and methods are suggested for converting semantic notions into a consistent system of images and characters. Ways of giving conceptual support in the digital transmission of such characters are examined. Principles based on work to date are summarised.

Ian Maun
University of Exeter Graduate School of Education
St. Luke's Campus EXETER EX1 2LU United Kingdom
Phone: 01626-772114; Email: I.C.Maun@ex.ac.uk

Received December 13, 2014; Revised February 17, 2015; Accepted March 4, 2015

Keywords: Real Character, lexicon, Basic English, composition, conceptual support.

1. 'Real Character' in the 21st Century: An Outline Summary of Maun (2013)

The history and principles of the concept of a 'Real Character' are outlined in some detail in Maun (2013). A 'Real Character' is a *pasigraphy*, that is to say, a system of writing designed only to be read, not to be pronounced. It aims to convey meaning directly to the mind without the use of conventional writing. Alphabetic, syllabic, and *abjad* scripts (e.g., Hebrew and Arabic) are all representations of the *sounds* of language, not of underlying meaning.

The term 'Real Character' (hereafter, RC) derives from the 17th century, when linguists and scientists believed it would be possible to classify all the elements of reality into a single system and devise symbols which would represent this reality directly as a form of writing or 'character.' A modern RC designed for use with computers and mobile phones would incorporate *icons*, *indices*, and *symbols*, in the Peircean sense. To create such characters it would be necessary to incorporate *semantic primes*, as explored by Wierzbicka and Goddard (e.g., Wierzbicka 1996, Goddard 2010) and to find *visual primes* which would convey meaning directly to the mind (Dondis 1973, Frutiger 1989).

In order to avoid the syntactic difficulties inherent in creating an auxiliary language such as RC, designed to be read by speakers of all languages, linear grammar is replaced with a format based on a T-bar. This arrangement enables the reader to choose the order of reading, e.g., Subject-Verb-Complement (S-V-C) for an English reader, S-C-V

for a Japanese one, and V-S-C for a Welsh one.

Since an RC of this type is designed to be used with digital devices, it is possible to incorporate computer affordances such as animation (e.g., to convey the idea of motion) and to present the text to the reader not as a block but in short or even individual units. This is known as Rapid Serial Visual Presentation (or RSVP) and readers can already achieve high reading speeds in conventional reading using this method.

2. Moving Forward

The present paper aims to devise principles for the selection of an appropriate lexicon for RC. Various possibilities are examined and the interim conclusions are framed as Guidelines for future work. A critique of some existing visual languages points out lessons and warnings for the future, and further principles for creating and reading semantically important characters are examined. As work on RC is at a developmental stage only, this paper may be regarded as a chapter of work in progress.

3. RC: An Auxiliary Language?

An RC such as has been outlined is not an international auxiliary language in the sense that Esperanto is. It is not a separately devised language which lies alongside a natural language or languages. It is, on the contrary, a semantic vehicle which will convey the meanings which a speaker or writer wishes to transmit, just as an alphabet is a

tool for transmitting sounds in written form. It is not necessary to learn a new set of words to replace those of one's native language. Rather, one must learn a set of visual and semantic conventions which will enable one to convey a message, in whatever language it is initially formulated, to a receiver who does not speak that language, but who understands the conventions of the pasigraphy.

This does not mean that an RC will be without problems of representation. Some languages convey concepts which are not easily translated into other languages, e.g., the Welsh word *hiraeth* and the Irish *hiraedd*, both of which convey a sense of nostalgia, homesickness, longing for the past and a sense of loss, and which cannot be translated into a single word in English. Similarly, English has no word for the French *fleuve* ('a river which flows into the sea'), the single word *river* covering both this concept and that of 'tributary river.' Likewise, some Australian aboriginal languages such as Yankunytjatjara have no generic word for *cloud*, only words for individual types of cloud (Goddard 1998). In German, to say *I am warm*, one has to say *Mir ist warm* ('To me is warm'), whereas in French one says *J'ai chaud* ('I have warm'). The three different constructions of English, French, and German must be expressible in one single way in RC.

Furthermore, semantics must be conveyed by syntax and a writer of Chinese has no morphological way of indicating the tense of a verb, which is normally marked only by context or adverbial modification. If that Chinese writer carries this convention over into RC, in a sentence unmarked in any way for tense, will it be possible for an English reader to place the meaning in the correct time-frame?

Such problems are, however, peripheral to the central concept of an RC and there are more pressing issues if such a system is ever to be devised. The way in which concepts may be pictured and visually

represented has already been outlined (Maun 2013) and it now remains to decide exactly where to begin with regard to the lexical content of the system. With which words or concepts does one start when devising such a system as is proposed here? Once chosen, how are such elements to be converted into a visual linguistic system? What examples of existing visual languages will give guidance or serve as warnings to the wise?

4. Audience, Purpose, Frequency

When devising a system for the teaching of *natural* languages to non-native speakers, it is necessary to consider a large number of factors with regard to the syllabus to be used. Devising the programme for an RC is in many ways similar. Munby's (1978) work is perhaps the most comprehensive examination of language syllabus design and takes into account a larger number of factors than could possibly be examined in a paper such as this.

His points to be considered include (a) the participant (b) the purposive domain (i.e., why the student is learning the language) (c) the setting (d) interaction (i.e., roles to be played) (e) instrumentality (spoken, written, face to face, etc.) (f) target level. As regards syllabus content, Munby specifies speech acts (e.g., explaining, advising) and a number of language forms, e.g., the language that would be required between a waiter and a customer. It should be noted that he does *not* give lists of nouns, verbs, adjectives, etc. to be learned. The lexis learnt will be dependent on precisely the above conditions and will not be decided in advance.

In developing an RC it is therefore necessary to look at the following related questions:

- 1) Who is to learn RC?
- 2) What is their purpose in learning RC?
- 3) With whom and how will learners be communicating?
- 4) What content are they to learn?

Given that RC is only at a developmental stage, it must be assumed that the learner is a generalist (and not a medical researcher or nuclear scientist) who will require a modern, non-technical, non-specialist vocabulary but who will need to deal with simple written texts (perhaps of the SMS type). Such a learner is likely to be communicating on everyday topics with similar general learners who do not necessarily speak the same language and who are possessed of an equally elementary RC vocabulary and syntactic ability. They will be using computers or mobile device screens, i.e., they will only be reading and writing. This suggests that the lexis first created for RC should consist of common or frequently-occurring words, provided that these can be created relatively easily from visual primes such as lines and basic shapes. The ways in which words and other semantic elements may be converted into such visual elements will be examined below.

(1) Guideline A:

The learner of RC needs a basic, modern, non-specialist vocabulary.

5. Elements for Inclusion

5.1. The Swadesh Lists

If an RC is to be able to fulfil its intended function of allowing speakers of any language to communicate in written form, a good lexical starting point might be ‘the Swadesh lists.’ These lists, which were developed by Morris Swadesh over many years (e.g., Swadesh 1950, 1971), catalogue words (or concepts) that are to be found in over 200 of the world’s languages. Appearing in several versions, the list was finally reduced in 1971 to 100 words. The list includes many function words, e.g., *this*, *that*, *here*, *there*, as well as many concrete lexical items, many of which relate to the natural world, e.g., *woman*, *egg*, *head*, together with verbs such as *eat*, *see*, and *hear*. That these concepts are found in so many languages might suggest that they are cultural or experiential universals. This, however, may not be the case (see below, section 5.2. Natural Semantic Metalanguage). Contemporary usage would also suggest that the vocabulary for near-universal cultural artefacts such as *mobile phone* and *computer* would now need to be added.

5.2. Natural Semantic Metalanguage

The work of Wierzbicka and Goddard (Natural Semantic Metalanguage) (e.g., Wierzbicka 1996, Goddard 2010) suggests that there are, in fact, better candidates for semantic universals than culturally recurring objects such as *egg*, *head*, or *phone*. Such semantic universals include, I, YOU, SOMEONE, SOMETHING, GOOD, BAD, BIG, SMALL, BE, DO, HAVE, SAY, SEE, NOT, MAYBE, THERE, and CAN. This latter list contains no concrete lexical nouns, but rather a set of semantic primes which fit together and from which propositions may be constructed, albeit circuitously. Thus ‘I have some ability in that field’ might be expressed as I CAN DO SOMETHING THERE. Sixteen of these semantic primes also occur in

the final Swadesh lists as free-standing words (*I, you, this, one, two, all, many, good, big, small, know, see, hear, say, die, not*) (my analysis). It was suggested in Maun (2013) that Goddard and Wierzbicka's primes should be adopted as elements from which RC characters might be created.

To adopt elements of NSM is not to say that words from the Swadesh lists should be ignored, but rather that such commonly occurring words should be adopted as lexical candidates alongside the more elemental Goddard and Wierzbicka semantic primes. It is, however, to be noted that *occurrence* in a language does not equate with *frequency* in that language, and words such as *louse* which are to be found in the Swadesh lists are unlikely to be frequently used in everyday communication, except, perhaps, among health workers. Their lack of frequency suggests that such words will not be particularly useful, unlike more semantically prominent words which will necessarily appear in a larger number of messages or conversations.

From these observations we begin to see criteria emerging which will assist in building a content programme for RC. These will include:

- 1) words which occur in many languages (from the Swadesh lists)
- 2) semantic primes (from Wierzbicka and Goddard)
- 3) frequently occurring and therefore prominent words

This last criterion, however, presents problems. A frequency count for one language will produce different results from a frequency count for another. Thus, in English the five most frequently occurring words are *the, be, of, and, and a*, but a frequency count of

French gives *de, la, et, le,* and *à*, i.e., *of, the* (masculine), *and, the* (feminine), and *to/at*. Both content and order differ between the languages but one notable fact begins to emerge from a perusal of such lists. The most frequently occurring words are *function* words, not lexical content words. This must be borne in mind when constructing a vocabulary for RC. Nouns will be needed for the expression of content in RC but priority must be given to function words. Some way of conveying such functional notions (which do not necessarily exist in every language) will need to be found. The first *noun* to appear in the English list is *year* in 60th position and in the French list, *gouvernement* ('government') in 98th position. Frequency is thus not an exact guide to the question of which lexical items should be included in any lexical ontology for RC, nor which will be most useful in everyday communication.

Note that an approach from frequency of occurrence, or, at least, an approach which partially takes frequency of occurrence into consideration, is the very opposite of the way in which the early 'language projectors,' e.g., Dalgarno (1661) and Wilkins (1668), attempted to create an RC. Their *modus operandi* was to divide the world into Aristotelian genera, differences, and species. At the ends of their taxonomic trees appeared actual words (or concepts). Semantic prominence was of no importance. How frequently such words appeared in any language was therefore not taken into consideration.

(2) Guideline B:

Use the Swadesh lists, NSM, and frequency lists.

5.3. Basic English

In 1930, the English writer and philosopher Charles Kay Ogden produced Basic English. This was not an international auxiliary language like Esperanto or Ido, both popular at the time, but a *modified natural language* (Large 1985: 160). 'Basic' was an acronym of British American Scientific International Commercial, a shorthand way for Ogden to communicate his intentions as to the uses and audiences for this modified form of English. He stated that Basic was 'a careful and systematic selection of 850 English words which will cover those needs of everyday life for which a vocabulary of 20,000 words is frequently employed' (ibid.: 163). By analysing the dictionary definitions of English words, Ogden discovered a limited number of frequently used words. As Pei (1958: 129) explains: 'By a careful statistical analysis of such words, coupled with the elimination of certain forms, Ogden concluded that it was possible to get along perfectly well in English with as few as 850 basic words, handled in normal English fashion, save for a restriction on the use of verb forms.'

Ogden's original word list consisted of 400 general nouns, 200 picturable objects, 100 adjectives, 50 adjectival opposites, and 100 'operators.' These latter consisted of 16 verbs (*be, come, do, get, have, give, go, keep, let, make, put, say, see, seem, send, and take,*) and other function words such as prepositions, articles, conjunctions, and adverbs. With regard to those words which were not 'operators,' Ogden had effectively found those which can be used to give dictionary definitions, e.g., *sort, kind, condition, act, process,* and which can also serve as combinable elements, rather like semantic primes, which can be linked together to provide the same meaning as more complex or abstruse words. By using such semantically central

elements, Ogden created economy of scale and allowed periphrasis to do the work of words which do not occur in this reduced language. To these elements were added words which were not so susceptible to semantic reduction, e.g., *moon* and *dry*.

It should be noted that 80% of the words in Swadesh's 100-word list also occur in Basic English (my analysis), showing that Ogden's system incorporates important semantic notions which are common to other languages. Note also that 39 elements in Wierzbicka and Goddard's Natural Semantic Metalanguage, including BE, DO, HAVE, SAY, and SEE occur as free-standing words in Basic English, as do functional elements such as WHEN, WHERE, HERE, NOW, and NOT (my analysis).

Basic English is not defined by the frequency of occurrence of the words themselves in English but by the frequency and prominence of the semantic elements that go to make up the meaning. The more a semantic element, e.g., *go*, can be used in the creation of legitimate English expressions, e.g., *go up* for *rise*, *go down* for *descend*, *go out* for *leave*, the more extensible it becomes. There is thus no need in Basic English for the verb *enter*, since this can be conveyed by *go in*, joining two fundamental units together. Similarly, *participate* is replaced by *take part* and *disembark* by *get off a ship*. Ogden himself demonstrates how terms can be substituted by translating passages of scientific English into their 'Basic' forms. A certain number of scientific terms have, naturally, been added to the core list for this purpose. Thus, *In the course of a systematic examination of the various products of carbonization ...* becomes in Basic English: *In the process of working on the different substances formed when coal is coked ...* (Large 1985: 168).

Words in the *General Basic English Dictionary* (Ogden 1940) are generally defined using Basic English words e.g., '**Wrist**—Join of

hand and arm,' but it sometimes becomes necessary to 'import' other non-Basic words to assist in the operations, e.g., *become*, as in: '**Go wrong**—*become* damaged, not in order.'

Charteris (1972) criticises Basic English, firstly, because it 'still leaves untouched the exceptional difficulty of an entirely irrational orthography and pronunciation' (p. 3), and, secondly, because:

[t]he Basic English word-list was merely taken as a convenient base to start from. Actually, it turned out, on analysis, to be a gentle deception, useful enough for getting the most out of the fewest English words, but achieving this by cunningly enlisting words which in English are loaded with multiple meanings, many of them with three, four, and even five ... *care, even, post, right, ring, and wind* [may serve] as examples (pp. 39-40).

Charteris's first point here need not concern us, since RC involves neither spelling nor pronunciation. The second point may serve as a warning, that to use the Basic English word-list as part of the foundation for the lexis of an RC may not in itself be sufficient. It may well be necessary to have recourse to other words or concepts in order to compensate for the lexical deficiencies in Ogden's system.

(3) Guideline C:

Use Basic English as the principal source of lexis.

5.4. Interglossa

Interglossa (1943) was the creation of the linguist, chemist, and philosopher Lancelot Hogben. Noting that the world was turning

ever more towards science and that the language of scientific discourse was founded principally upon Latin and Greek roots, he devised an isolating language based upon Classical words and radicals.

Hogben followed Ogden in using combinable verbs, which he called *verboids*, as follows:

(4) Hogben's verboids

<i>habe</i> (have)	<i>reacte</i> (react)
<i>gene</i> (get)	<i>esthe</i> (feel)
<i>date</i> (give)	<i>vise</i> (see)
<i>perde</i> (lose)	<i>dicte</i> (say)
<i>tene</i> (keep, conserve)	<i>acouste</i> (hear)
<i>detecte</i> (find, discover)	<i>kine</i> (move [intrans.])
<i>acte</i> (do, perform, carry out)	<i>mote</i> (move [trans.])
<i>tracte ... apo</i> (take ... away)	<i>facte</i> (make, construct)
<i>balle</i> (dispatch, send, cast, throw)	
<i>stimule</i> (excite, evoke, stimulate, call forth)	

These verboids correspond to a number of Ogden's operators, but not all—*perde*, *detecte*, *stimule*, *reacte*, *esthe*, *kine*, *mote*, and *acouste* have no exact equivalent in Basic English, although nouns such as *discovery*, *hearing*, and *loss* are to be found in that word-list. Hogben's verboids can be used literally but when combined with certain nouns (which Hogben calls *substantives*) they form a kind of periphrasis which can be used to express notions for which no actual verb exists in Interglossa. Thus *to warm* is expressed as *date thermo*, i.e., to give heat, *to cool* (intrans.), *perde thermo*, i.e., to lose heat.

There are, of course, certain ontological problems with such a

system, which presupposes that heat can be separated from the processes of heating and cooling and that the processes of heating or cooling involve the transfer of a property from one thing to another in the same way that an object is physically handed to or removed from a person, i.e., by giving or taking. However, this metaphorical way of expressing these processes is probably adequate for everyday purposes of communication, even if it might fail a strict theoretical scientific test.

(5) Guideline D:

Incorporate the outstanding Interglossa verboids into the BE list.

6. Affixation

In linguistics the term 'affixation' has two distinct senses. In polysynthetic languages such as Nootka (Nuu-chah-nulth), an entire message can be coded in one word whose root is a verb to which other semantic and syntactic elements are attached by prefixation, suffixation, or both (Sapir 1921). These elements may correspond to distinct words or morphemes in English and other languages. It is not however with this process that we shall be concerned here, but with the second sense of 'affixation.'

This second sense is the process whereby a root or radical is modified by prefixation, suffixation, or infixation to give a more precise or specific meaning to a root, e.g., *to date* becoming *to pre-date*, or to change a root into another grammatical category, thereby changing both its meaning and its function, e.g., *pretty* (adjective) becomes *prettify* (verb). In English, French, and many other languages, prefixation does not change the category of the root but

merely adjusts or focuses its meaning. It is thus normally a productive process which allows the creation of a large stock of words from a set of roots and affixes. (Some prefixes have a very limited application, e.g., in English, the prefix *areo-* is reserved exclusively for the meaning *relating to the planet Mars*.)

Suffixation on the other hand changes the syntactic category of a word. This latter process is known as *derivation*. Thus the French root *surveill-* expresses the notion of ‘supervise.’ By adding the suffix *-er*, we produce a verb, *surveiller* (‘to supervise’) but by adding *-ant* we produce the present participle *surveillant* (‘supervising’) or the noun *un surveillant* (‘a supervisor’). In Arabic, suffixation can also change the category, as in English: *jamaal* (n., ‘beauty’) becomes *jameel* (adj., ‘beautiful’). In that language, *prefixation* can also be used to change the category while retaining the central semantic concept: *tajmeel* (v., ‘to beautify’). In each case, the affix identifies an important semantic or syntactic difference. Note, too, that the core of a word can express central semantic concepts in Arabic using consonants alone. Thus *k-t-b* gives *keteb* (‘book’), while *maktub* means ‘office’ (i.e., place of books).

Affixes thus focus on semantically prominent notions. In English, prefixes can express the notion of:

- (6) TIME and ORDER (PREnuptial, ANTEnatal, POST-war),
- PLACE (SUPERstructure, SUBmarine, INTERNATIONAL),
- NEGATION (UNnecessary, INSane, ILliterate),
- REVERSAL or PRIVATION (UNdo, DEcentralise, DISconnect),
- DEGREE or SIZE (ARCHbishop, MEGAton, SUBstandard),
- ATTITUDE (COoperate, COUNTERact, ANTibiotic, PRO-Europe).

Prefixes can also convey:

(7) PEJORATIVE meaning: MISlead, MALTreat, PSEUDO-intellectual.

Suffixes in English can express, *inter alia*, the notion of

(8) PERSON (prankSTER, privateER, teachER, professOR, merchANT),

SIZE (pigLET, cigarETTE),

STATUS OR DOMAIN (brotherHOOD, kinSHIP)

(Examples from Quirk et al. 1973)

Esperanto makes wide-ranging and systematic use of affixation, suffixes outnumbering prefixes by a good margin. Some suffixes serve to show the syntactic function of a word. Thus *-o* is the suffix for nouns, *-a* for adjectives, and *-i* for verbs. Because of this feature, other suffixes can become infixes as they must precede the grammatical suffix, e.g., the suffix for 'large' is *-eg*. If added to *domo* (house) to create the Esperanto word for *mansion* (i.e., *large house*), the suffix *-eg* becomes an infix before the final *-o*. Thus, *domego*.

This style of affixation is not confined to nouns. Thus, since the suffix *-et* is used to create a diminutive form, the adjective *varma* ('warm') can become *varmeta* ('lukewarm'), and the verbal form *dormas* ('he sleeps') can become *dormetas* ('he dozes,' i.e., 'he has a little sleep'). This transcategorical phenomenon exists in English as, well, of course. The prefix *pre-* ('before') can be used with nouns (*pre-war*), with verbs (*to prefix*) and with adjectives (*pre-determined*). The same is true of many other prefixes and suffixes. Since such meaning affixes are also trans-categorical in many auxiliary languages, it might be better to refer to them as *semantic meta-units*.

Hogben (1943: 98) points out that 'Interglossa has no lifeless prefixes.' What he means by this is that no element prefixing another word is itself a bound morpheme, such as *pre-* or *post-*. All elements

used as prefixes in Interglossa are free morphemes which can also have other functions such as prepositions. Thus *ante* means ‘before’ when used as a preposition and ‘pre-’ when used as a prefix with a substantive. This economy of scale thus reduces the number of elements needed to create compound words.

Basic English also uses this principle. There are no prefixes or suffixes in the 850 word list. Thus the verb *to misconceive* is expressed by *to get a **wrong** idea of*. There is no separate prefix corresponding to the idea of ‘wrong,’ which is a free-standing word in the vocabulary list. Similarly, negative prefixes such as *il-* (*illiterate*), *im-* (*immediate*, *impossible*), *in-* (*inadequate*) and, *ir-* (*irresistible*) could all be expressed by prefixing the positive word with *not*, a basic word from the vocabulary list, e.g., *illiterate* means *not able to do reading*, using nothing but Basic words. We shall return to this point when we examine ways of converting vocabulary into symbols in RC, using such symbols as the logical negative ‘¬’.

It is evident that such meta-units may have wide application in RC, since they carry important meanings. It would be necessary to reduce the sense of such units to another semiotic form, either iconic or symbolic, which could be attached to characters to alter their sense. Note that this would not always be necessary, since the syntactic function of words is in part shown by their place-value in RC. Thus the symbol for ‘see’ (perhaps \odot) would mean *to see/sees/saw* etc. if placed in predicator position, but would mean *seeing/sight/view/visibility*, etc. if placed in subject or complement position. The exact choice of meaning in these cases would be determined by some other symbolic modification, yet to be determined, e.g., the eye symbol would need to be combined with a symbol for *possible* in order to express the notion of *visibility*. See below, section 9.3. Composition of Characters.

The creation of a list of meanings which can be expressed through meta-units might, of course, be no easy task. According to Hanks (1992), Urdang (1982, 1984) has identified no less than 4,400 affixes in English. Many of these will be specific to certain areas of knowledge, e.g., medicine, science, engineering. It would therefore be necessary to identify a core set of meanings expressible by meta-units which could be used in a preliminary, everyday version of RC. Languages other than English may express such central concepts by different syntactic or semantic means. Affixation is by no means a universal phenomenon. Isolating languages, by definition, do not use such meta-units, but this does not mean that these languages are incapable of expressing a given meaning. They just do it differently, usually by means of another word or construction. RC, being a potentially universal means of expressing one's own language in a written form, must ensure that meanings are expressible without favouring any particular language or language family.

So how can we arrive at a list of semantically prominent meta-units which can be applied to various syntactic categories? Using Jacobs (1947), an analysis of major artificial languages of the 20th century (Esperanto, Ido, Novial, Occidental, and Interlingua) reveals that most of the semantic meta-units used in these languages are, in fact, covered by Basic English (BE) words e.g., Esperanto *-isto* = BE *person*; Interlingua *pseudo-* = BE *false*; Ido *-aj-* = BE *bad*; Occidental *-tá* = BE *quality*; Novial *-ilo* = BE *instrument*. The meanings of meta-units in these languages which are *not* covered by BE elements can be made from periphrasis, e.g., Esperanto *-iĝ-* (*become*) = BE *change* + *to*; Ido *-ig-* (*make, render, transform into*) = BE *cause* + *change* + *to*; Novial *-endi* (*must be, worthy*) = BE *necessary, give* + *reward*. Lest it be thought that more modern artificial auxiliary languages are receiving no attention here, it is

necessary to note that the core vocabulary of Lojban is that of Basic English, although the concepts thus covered use entirely different linguistic forms.

It is, then, apparent that Basic English not only provides a powerful list of words which correspond in part to the Swadesh lists, to parts of Interglossa and to elements of NSM, but also provides a highly flexible list of semantic elements which can be ‘glued’ to other words to extend or qualify their root meanings in the same manner as that employed by major artificial languages.

(9) Guideline E:

Establish and use semantic meta-units where possible.

7. A Note on Vocabulary Size

By adding together the 850 words of Basic English, the additional *verboids* from Interglossa, and the semantic primes of Wierzbicka and Goddard, and by adding in the combinatorial possibilities of a number of meta-units, one might conceivably arrive at a total lexicon of a few thousand vocabulary items. Many of these would, of course, be periphrastic rather than single words. Just how useful would such a limited vocabulary be?

The French equivalent of Basic English, *le français fondamental*, has only 1,475 words at *premier degré* (Level 1) and 1,609 at *2e degré* (Level 2). This is a miniscule lexicon. The vocabulary of the visual system known as Picture Communication Symbols consists of about 10,000 pictographs (Ting-Ju & Biggs 2006). Even this is a small vocabulary for an adult. Pinker (1995) notes that a native speaker can understand as many as 40,000 words. (Coincidentally

and interestingly, *The General Basic Dictionary* (Ogden 1940) gives 40,000 senses of 20,000 words.) Many of these, of course, are synonyms, or near-synonyms, e.g., *likely* and *probable*, *relative* and *relation*, *tough* and *resilient*. A glance at a thesaurus reveals that many words merely express subtler shades of other words which could be expressed by the original word and a modifier, e.g., *gorgeous* = *very* + *beautiful*.

Research into second-language learning reveals some useful figures. Tseng & Schmitt (2008) summarise findings from a range of research on vocabulary learning for second-language learners with the following figures, where the L2 is English.

- 2,000-3,000 word-families are needed for basic everyday conversation (chat).
- 3,000 word families are needed to begin reading authentic texts.
- 5,000-9,000 word families are needed to independently read authentic texts.
- 10,000 word families, a wide vocabulary, are needed to allow most language use.

The notion of word-families (a head word, its inflected forms, and its closely related derived forms, e.g., *like*, *likely*, *unlikely*, *unlikelyhood*) links to the idea outlined above of root words and additional semantic meta-units.

Given the developmental stage of RC at present and the targeted learner ('a generalist'), we may therefore not need a greatly expanded version of our linked 'Basic English + Interglossa verboids + NSM + meta-units.' As with many things, 'the proof of the pudding is in the eating' and we shall have to await the results of

further developments of research into RC and its applications to digital devices.

(10) Guideline F:

Develop the means initially to produce a vocabulary of 2,000-3,000 word-families.

8. Visual Languages

8.1. Hanks (1982): Eh May Ghee Chah

Eh May Ghee Chah, the ‘universal second language’ devised by E. J. Hanks (Hanks 1982) is not, strictly speaking a visual language and is therefore dealt with here before true pictographic systems. Composed of straight lines, right angles, and dots, its 56 characters represent syllables. When combined in topic matrices, a group of these characters represent a word or concept, which may be found either by reference to the matrices or to topic lists. The vocabulary is mainly drawn from Basic English and Roget’s *Thesaurus*. Hanks estimates the total word-count to be about 5,000.

Hanks uses two characters at the beginning of every word in his syllabic language. The result is that every word from a given category resembles every other. Given that all his characters consist of lines and dots, some words are virtually identical to each other apart from, say, a single dot. Thus the word for *cathedral* is virtually indistinguishable from the word for *pew*, both falling in the ‘ecclesiastical’ category. This is a serious disadvantage for the learner.

Although Eh May Ghee Chah resembles a visual language, it is more closely related to the category-based Real Characters of Wilkins and Dalgarno. To use the language would require constant reference to lists and matrices or a memory of almost infinite capacity.

(11) Guideline G:

Avoid the over-use of categorical indicators.

8.2. Charteris (1972): Paleneo

Charteris (1972) originally developed his symbolic language Paleneo out of the shorthand that he devised for his own purposes when writing his crime novels. To give this system structure, he classifies a number of basic words under topic headings, e.g., Numbers, Pronouns, Directions, Transport, Time, and Places. Within each category, he uses a basic symbol or symbols from which to develop other characters, e.g., arrows under Direction, a figure resembling an hour-glass under Time, a simple square (□) for Place. Not all symbols are developed from the prime, however, e.g., under Place, 'room' is given as □□, which is taken from the Egyptian hieroglyph for 'house,' not from the basic Place symbol, □. Charteris is therefore inconsistent in his use of categorical markers.

Against this criticism must be placed the fact that he avoids the pitfall of marking all words in a category with the same symbol, thereby making all words in that category resemble each other, the trap into which Hankes (above) falls all too easily. Perhaps such categorical information should be available elsewhere, as outlined below under section 10. Conceptual Support.

(12) Guideline H:

Be consistent in design but avoid total resemblance.

8.3. Randič (2010): Nobel

Nobel (2010) is a visual language of some 600 words devised by the Czech researcher Milan Randič. He begins by creating 60 visual characters, which he calls *symbols*, and which he describes as ‘obvious.’ It is not made clear to whom the symbols will be ‘obvious,’ but as many are iconic, their meaning is relatively transparent. Other characters are clearly *symbols*, in the Peircean sense, i.e., arbitrary characters without isomorphism, and their composition requires some explanation before they become ‘obvious.’ This is, indeed, a problem which plagues artificial visual languages—it is often easy to see a symbol’s meaning once it has been explained but it is not always easy to see a meaning without some explanation. Nor can one pre-determine the form of a symbol from the meaning of a word with absolute certainty. Blissymbolics is a notable case in point. See Okrent (2009: 166) for discussion.

Nobel uses a number of principles in the construction of characters, such as repetition for some plurals and overlaying to convey the notion of ‘combined ideas.’ Much thought has gone into the creation of this visual language and many of its principles could be adopted in the creation of a contemporary digital Real Character.

However, although published in 2010, Randič’s book has no mention of digital devices such as computers or mobile phones. Indeed, with its emphasis on drawing, Nobel could be simply a more advanced and more developed version of Charteris’s Palneo. Furthermore, Eco (1995) dismisses an early version of Nobel as being among ‘lexical codes without any grammatical content.’

Indeed, as far as one can judge from Randič's translations of Chinese sayings into Nobel, the word order is simply that of English. In its favour, however, the book has the advantage of being incremental in its approach—the learner is first exposed to a number of basic characters and more developed or complex symbols are built up from these, with the author keeping a running total of the 'words' thus accumulated as the book progresses.

(13) Guideline I:

Be incremental in approach.

9. On Converting Words into Symbols

9.1. Primes and Conventions

The choice of the vocabulary of Basic English as a provisional starting point for the lexis of RC has a number of advantages, one of which is that Ogden lists 200 picturable objects (many of which occur in the Swadesh lists). RC will require that characters for these picturable objects be *iconic* (isomorphic with the referent), and so some linear representation will be chosen. This may be based on a photograph but since the style of photographs can vary enormously, it may be necessary to adopt a consistent linear style which borders on the cartoon rather than a detailed image.

McCloud (1993: 49) makes the point that pictures are *received* information (they can be understood instantaneously) whereas writing is *perceived* information: 'It takes time and specialized knowledge to decode the abstract symbols of language.' Nevertheless, some degree of conventionalisation of pictures will be

necessary, i.e., a stereotypical image will be required rather than the image of particular person or thing. Thus, for the notion of *footballer*, a typical footballer would be drawn, rather than using a photo of, say, David Beckham, as this latter could be misinterpreted as ‘David Beckham’ rather than ‘footballer.’

It will furthermore be necessary to adopt certain conventions in the design of such iconic characters, e.g., it might be agreed that all human beings are drawn as from the front, whereas four-legged animals might be drawn from the side to show up the differences between various types, and insects could be drawn as from above to show wings, patterning, appendages, etc. This corresponds with the fact that we normally view humans and large animals from our own vertical perspective, whereas insects are recognisable principally from an overhead view.

While concrete beings and objects present little difficulty in terms of iconic image design, abstract concepts, and structural function words present their own problems. Concepts such as these can only be expressed in a symbolic manner, i.e., they can have only a form conventionally agreed by its community of practice, e.g., musical notation, chemical symbols. The characters for such words or concepts in RC will be formed from visual primes which themselves carry meaning, e.g., straight lines (rigidity, firmness, etc.), curved lines (flexibility, softness, etc.), and geometric shapes (squares, rectangles, circles, etc.). Each of these latter may be associated with certain meanings, e.g., the square may be seen as representing solidity, balance, firmness, etc., the circle, wholeness, completion etc. See Maun (2013) for further elaboration.

9.2. Haag (1902): *Outline of a Logically-Based Graphic Language* [Translated title]

The work of Haag (1902) may provide a basis for such design work. Haag's starting point is the writer as centre-point for the communicative process:

The notion of 'space' forms the basis for the organisation of our concepts; spatial concepts must therefore become the elements used in the logical construction of the expression of thought; because of their close relationship with the visual faculty, they provide the most direct means of symbolisation. ... The spatial behaviour of two things in relation to each other forms the natural basis of visualisation. Basic predication, synonymous with the relationship to the underlying idea of the second thing, is derived from the first. *Horizontal* and *vertical*, *distant* and *enclosed* provide the four visual primitives. To each correspond two opposite forms, consisting of contrary concepts, *in front of* and *behind*, *above* and *below*, *near* and *far*, *inside* and *outside*.

Closely related to spatial concepts are those of *number* and *measure*, and, in part, these derive directly from notions of touch, and in part are transferred from [the concepts of] space. *Much* and *little* correspond to *above* and *below*, with allowance being made for the underlying concept of an intermediate measure. The extreme limits appear to be *all* and *nothing*. *Near* and *far* correspond to *present* and *absent*, all 6 of these being used in their absolute sense. *Similar* and *dissimilar* (*like* and *unlike*) closely resemble *near* and *far*, in their relative senses. Independent of the above concepts are

definite and indefinite, whole and partial (collective and individual in relation to many) (pp. 2 & 4-5, my translation).

Haag takes a number of visual primes in the form of lines and dots to create semantic primes and extends their use to metaphorical purpose, thereby forming a conceptual bridge between the concrete and the abstract. This conforms with Jackendoff's (2012) notion of a conceptual structure bridging the mental gap between the visual plane (seeing an object) and finding a 'handle' by which to identify it (the pronunciation or phonological representation of the word attached to the object). Because of the close conceptual relationship between 'near' and 'similar,' and 'far' and 'dissimilar,' the visual symbols for both the physical and conceptual relationship may serve as a useful basis in constructing RC symbols. In an RC, *depicting* this relationship replaces the phonological 'handle.'

Haag is furthermore able to extend the use of his basic concepts beyond simple adjectival use to the fields of adverbs and conjunctions:

It was [also] necessary to undertake similar work to define the underlying concepts for adverbs, conjunctions, and structural symbols which express the predication of two ideas (whole sentences in separate clauses). The adverbs *and, likewise, also, too, not only ... but also* all come under the concept of *being present*; the adverbs *only not, neither ... nor*, and the conjunctions *except that, without* come under the concept of *being distant, missing*; the adverbs *likewise, also, suchlike* come under *agreement*; the adverbs *but, on the contrary, but, however* and the conjunctions *against which, whereas* come under *disagreement*. They mean the same as

the prepositions *with*, *without*, *like*, *unlike*; for *or*, the classification *uncertain* matches the concept of *exchange*. In exactly the same way as we have treated logical predicates, for temporal ones, we recognise the concepts *precede*, *follow*, *occur*, *coincide*, *follow immediately*, in which are concealed the prepositions *before*, *after*, *at the time of*, *in*, *within*, *immediately after*; for causal predicates, *derive from*, *aim at*, *pass through*, *depend*, expressed through the prepositions *out of*, *to*, *through* (pp. 26-27, my translation).

Haag's prepositions can also become conjunctions: *out of* becomes *because*, i.e., one situation emerges from another, the first situation being the cause of the second. A symbol for *because* should therefore show emergence if it is to be iconic, e.g., \ominus , which is possibly preferable to \because , the mathematical symbol for *because*. Presumably *therefore* is represented by a sign showing the meaning of *into*, because 'X, therefore Y' means 'X leads into Y.'

Since Haag is writing in the European tradition, his graphic language is read from left to right. Thus *in front of* is expressed by a dot followed by a vertical line, *behind* by the vertical line preceding the dot. By analogy, these can be used temporally to express *before* and *after*. This leads to the possibility of expressing verbs such as *precede* and *follow* by using these same symbols.

Given that RC is to be read in the natural linguistic order of the reader, e.g., English: S-V-C, Japanese: S-C-V, Haag's basic left-to-right symbols may pose some problems for the system. Linguistically careless advertisers have also run into this problem, e.g., English-speaking designers advertising pain-killers with 'Before' and 'After' pictures have failed to realise that Arabic readers read from right to left. This produces comic and possibly unprofitable results! The

problem of structural ordering within symbols and the direction in which they are to be read will be discussed in the next section.

(14) Guideline J:

Use Haag's approach to the analysis of connections between concepts.

9.3. Composition of Characters

In Maun (2013), the question of syntactic ordering within the sentence in RC was addressed, with the functionally-based T-bar being proposed, as outlined in the introduction to the present paper. Furthermore, it was proposed that adjectives should be placed below the noun in order to solve the problem of pre- or post-positioning. A similar approach must now be taken to the question of order *within* characters.

In order to achieve automaticity of recognition, of course, a character must not occupy too much space, or it begins to resemble a sentence. This is one of the criticisms which may be levelled against Blissymbolics, which can use four or five symbols on a line to represent some notions, e.g., *decoration, ornament*.

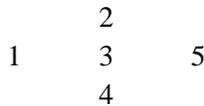
In RC, if we are dealing with a simple glyph, of a type such as the Blissymbolics icon for *house*, , there is little problem. Once we start dealing with compounds, however, questions of position and order are immediately raised. Take, for example, the concept of *invisibility*. This consists of at least four elements: [NOT - POSSIBLE - SEE] + ABSTRACT QUALITY. It will thus have to be related in RC form to other concepts such as *impossibility, illegibility, and unlikelihood*, in which similar components are semantically present, even if their surface forms are different. It will, therefore, be advantageous if such

elements occur regularly in the same position.

Examples from natural languages may serve as guides to the way in which such a question may be solved. In Chinese, if the character for *grass* is used in a compound, it always occurs at the top of the compound character (Scurfield & Lianyi 2003). Similarly, 'the semantic radical meaning *hand* or *actions related to the hand* must appear on the left side of a character' (Feldman & Siok 1997: 776).

To take another language—'Compounds' in historic Maya texts are groups of glyphs wedged tightly against one another. Houston (1989: 33) explains how such glyphs are grouped and read:

Reading order within the compound is consistent but by no means rigid. The affixes to the left or top—known as 'prefixes'—were read first. Those to the bottom and right—the 'postfixes'—came last. The centre, place usually occupied by a main sign, was read between. This diagram gives an idea of the sequence:



It is here we get our first glimmering of reading order: left to right and top to bottom.

If we put these two principles together, we arrive at a possibility for the organisation of character-elements in RC:

- 1) Elements with given meanings occur in fixed positions within a character.
- 2) Elements are read in a given order.

Thus, if we take two symbols already mentioned, the logical sign for negativity, ‘¬’, and the ‘eye’ symbol, , and add a temporary abstract symbol in the form of ‘IO’ for *possible*, we could link the two abstract elements at the top (NOT + POSSIBLE) above the lexical element (SEE). This would give ‘invisible’:

(15) ¬ IO


In Egyptian hieroglyphs, we find the symbol showing a scroll used to mean ‘book’ or ‘abstract quality.’ This, or a similar symbol such as , could be added beneath our partially constructed glyph, to give the following for the abstract notion of *invisibility*:

(16) ¬ IO



Following these principles, *illegibility* might appear thus:

(17) ¬ IO



Once such principles are established, it becomes relatively simple to work out the meaning of a glyph such as:

(18) ¬ IO



At this point, it becomes necessary to solve the difficult problem of reading order. T-bar syntax for sentences can be read in any order, depending on one's native language. We cannot, however, use the T-bar within a glyph, making it possible to read a character in any order, as the potential meaning and functional elements which are available do not fall neatly into a tripartite Subject-Verb-Complement format. While a particular order of assembly for elements such as TIME, POSSIBILITY, PLACE, NEGATIVITY, PERSON WHO ..., etc. remains to be determined, we can adopt the Maya reading principle: *Within* the character, whatever one's native language, always read top to bottom, left to right. While this may seem arbitrary, any other order would be equally arbitrary and nothing is gained from moving away from a principle already used in an attested visual language.

(19) Guideline K:

Establish formational and reading parameters for characters.

9.4. Correlatives

Esperanto (copied by Ido) has a system of joining related concepts together. The function words 'which/what,' 'that,' 'some,' 'no,' and 'each/every' can be combined with the following notions: *one, thing, kind, place, way, reason, time, quantity, one's* [sic]. Each of the function words begins with a particular letter, thus:

(20) which/what	=	<i>k-</i>
that	=	<i>t-</i>
some	=	<i>i-</i>
no	=	<i>n-</i>
each/every	=	<i>c-</i>

Thus ‘which one’ = *kiu*, ‘which thing’ = *kio*, ‘which kind’ = *kia*, ‘which place’ = *kie*, etc. In parallel, ‘that one’ = *tiu*, ‘that thing’ = *tio*, ‘that kind’ = *tia*, and ‘that place’ = *tie*. Semantic elements are realised systematically in a clear and consistent way.

Pankhurst (1927) criticises this system, not for any inherent flaw in its thinking and combinatorial possibilities, but for the difficulty of pronouncing certain combinations. This is a poor criticism in the light of the undoubted advantages that this system has over natural languages and one which will not concern us when dealing with a pasigraphy such as RC. Compare the lack of connection between the French questions *lequel/laquelle?* (‘which one?’) and their answers *celui-là* or *celle-là* (‘that one,’ according to gender), or the German equivalents *welcher/welche/welches?* and their answers *derjenige/diejenige/dasjenige*. While syntactic initial elements in the latter case show some connection to the question, there is no link in the root form between question and answer. Esperanto correlatives score well in this respect and offer a promising way of creating linked items in the RC lexicon.

It might thus be possible to develop symbols, e.g., for *Which one?* In which question and answer resemble each other:

(21) Question: $\left. \begin{array}{c} \vdash \\ ? \\ \vdash \end{array} \right\} = \text{Which one? (i.e., This one or that one?)}$

Answer 1: $\left. \begin{array}{c} \vdash \\ \vdash \end{array} \right\} = \text{This one}$

Answer 2: $\left. \begin{array}{c} \vdash \\ \vdash \end{array} \right\} = \text{That one}$

(22) Guideline L:

Ensure that correlatives resemble each other, e.g., in question and answer.

10. Conceptual Support

Jackendoff (2012) makes the point that, in the visual field, if an element of a picture is missing, our *gestalt*-oriented mind completes the picture or restores the missing part, e.g., if we see an image of the rear end of a cat disappearing behind a bookcase, we 'restore' the front part automatically. We know that there is a whole cat there. In other words, there is more conceptual information available to us than the simple 'handle' by which we grasp the image. The same is true of words. While we can understand the word *oak* as referring to a tree, even some particular tree within our field of vision, we actually know far more about this referent than its visible shape. We have encyclopaedic (or conceptual) information available to us. A sentence such as *A holm-oak is exceptional because it's evergreen* is immediately interpretable, since we have subconsciously-stored information available to us that oaks are normally deciduous trees, not evergreens. No such information is available within the surface pronunciation 'oak.' It is this underlying conceptual information that enables us to make logical inferences and deductions.

In RC, *symbols* have fewer obvious information-elements than do *icons*, since they represent abstract concepts which are, by definition, not directly picturable. The nearest that we can get to 'picturing' such concepts in RC is through the use of such means as 'time-as-space' metaphors, as in Haag, the use of colour to distinguish them from literal meaning and the use of straight lines for 'hard' concepts and curved lines for 'soft' ones and harmonious or unharmonious shapes.

Esquisabel (2012: 10) notes that for the 17th-century philosopher Joachim Jungius, 'a symbolic notion consists in the consideration of a thing by means of the use of a name, without thinking of the aggregate of concepts that constitute the definition of the referred

thing. We act in this way—says Jungius—in order to think quickly and shortly, knowing that we can recall the constituting concepts at will.’ Since symbols are lacking in such ‘constituting concepts’ (as are some icons), it may be necessary in RC to add a further layer of information, perhaps in the form of animation or features derived from componential analysis. Such features could be added to the surface symbol or at a different digital layer which could be called up on demand on a device such as a mobile phone. Thus, if we choose  as a metaphorical symbol for *thought*, we may need to have [+ ] and [+ ] as underlying support, or [+ HUMAN] and [+ BRAIN], if conceptual support is chosen as being given in a particular language. The cloud symbol is thus shorthand for a bundle of features (as is any word or symbol), which may be expanded to enlarge or clarify information. This may be particularly important in the case of *indices*, where an iconic picture of smoke might be linked to underlying notions of ‘fire.’

(23) Guideline M:

Establish conceptual support at an underlying digital level.

11. Applications

If the foregoing reasoning is valid, then it should be possible to demonstrate that sentences in various languages can be re-cast into a form which is expressible in the vocabulary composed of Basic English, Interglossa verboids, and NSM primes (all ultimately converted into RC characters, of course). Furthermore, it may be shown that ‘screen level 1’ (SL1) representations of RC on

computers and mobile devices may be reinforced by an underlying 'conceptual support' level, as discussed above, which can be called up as 'screen level 2' (SL2) to clarify or amplify the meaning.

Of course, the actual characters for an RC will require an entirely new font (as did Wilkins's 1668 version) and we shall have to be satisfied for the present with existing characters which bear some resemblance to future representations, or which at least may be allowed as arbitrary symbols, e.g., 'IO' for 'possible,' as above.

An example from an obscure language shows that where the semantics does not exactly overlap with English, concepts may still be matched by the use of semantic primes drawn from our three sources. Everett (2012) discusses Piranhã (pronounced 'pee-da-HAN'), a language of Brazil with about 150 speakers. From Everett's examples, it may be seen that this language has in its vocabulary the nouns *animal*, *baby*, *bed*, *boy*, *father*, *man*, *mother*, *snake*, *son*, *woman*, and *word* as well as the verb *bite* and the adverb *almost*, all of which match well with terms in Basic English. Very close equivalents are *speak* (BE *say*), *want* (BE *desire*), *pretty* (BE *beautiful*), and *people* (BE *persons*).

On the other hand, Piranhã has only four colour terms **biísai** ('red'), **koobiai** ('white'), **kopaíai** ('black'), and **xahoasai** ('green/blue'). These could easily be represented in RC either by the colours themselves at SL1 (as discussed in Maun 2013), or, at SL2, by BE words in RC form. It must be noted, however, that the Piranhã terms are not, in fact, words, but phrases. Thus **biísai** is a phrase meaning *similar to blood*. This may be expressed in RC at SL2 as [LIKE + BLOOD], using BE terms. Piranhã also has verbs like *eat* and *kill*, which have no direct equivalent in BE. Iconic representations at SL1 might not necessarily be clear but SL2 representations could use symbolic forms of BE synonyms in RC form for clarification, e.g.,

[TAKE + FOOD] and [BE + CAUSE + OF + DEATH + OF]. Such a solution would, of course, require empirical verification with native speakers.

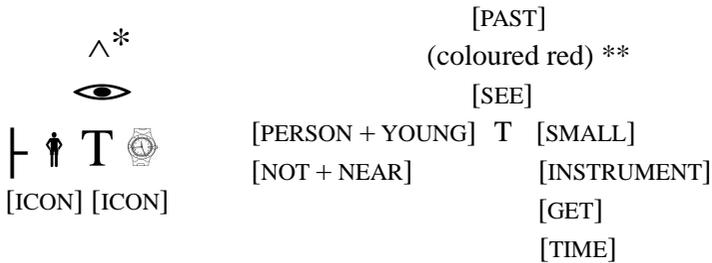
Time phrases in Piranhã also require some attention. While the language has terms for *night*, *day*, and *noon*, these, like the colour terms, are expressed as phrases with metaphorical import. Thus, *day* is **hoa**, which literally means *fire*. *Night* is **xahoái**, which literally means *be at fire*. So, *night* might be symbolically expressed at SL1 by a character similar to the Blissymbolics one for *night*, , but might require conceptual support for a Piranhã speaker by a representation at SL2 of the Basic English terms [BE + AT + FLAME] in symbolic form. Other temporal expressions in Piranhã such as *noon*, *sunset*, and *sunrise* would require similar support in the form of periphrasis.

Let us take another, unrelated language—Japanese. In the example below, where an icon is used in the RC representation, as in the case of something concrete and visible, we can represent this here at ‘screen level’ by the use of an image and the temporary indication [ICON] below it. This indication would not appear on screen. At conceptual level, componential analysis and/or animation, colour, etc. will provide support.

(24) Sono kodomoga tokeio mita.
 that child-SUBJ watch-DIR OBJ saw
 'That child saw the watch.' (Demers & Farmer 1986)

(25) **Screen level 1 (SL1)**

Conceptual level (SL2)



* The symbols \wedge , $-$, and \vee are used to mark past, present, and future on the verb. The 'up-down' directionality avoids left-right or right-left reading. Such symbols might be optional. A Chinese speaker would not use them, tense not being indicated in Chinese.

**A 'traffic light' system of red, amber, and green is used to indicate past, present, and future.

Note that the semantic elements used here in the componential analysis at conceptual support level are either Basic English words, an Interglossa verboid, or an NSM prime. These may ultimately be replaced at this level by recognisable symbols such as the negative '¬'. Similarly, [NOT + NEAR], i.e., 'that,' could be represented by Haag's | | ('far, distant'), as opposed to ||, 'near,' i.e., 'this.' The

present SL1 symbol ‘ \vdash ’ is taken from the temporary system of correlatives discussed above.

A Japanese reader would read this sentence in the order S-C-V, whereas an English reader would use S-V-C. The syntactic suffixes *-ga* (subject) and *-o* (object) no longer appear, of course, and their function is indicated by *place-value*, as discussed in Maun (2013). Conceptual level support could be called up in Japanese script or further symbolic representation. Remember, the Japanese sender may wish to check that the screen level message in RC does in fact mean what he/she wants it to say!

(Note that a Mandarin reader could read both the SL1 and SL2 forms not as S-V-C, but as Topic-Comment: ‘Watch - that child - see [it],’ which would conform to the syntactic preferences of that language.)

Basic English may thus form (a) a core body of words which are found in most languages, and which are shown at SL1, and (b) a set of semantic units (later converted into symbolic form) which may be employed on a different screen of the device (SL2) to clarify or amplify the meaning according to the reader/writer’s requirements.

(26) Guideline N:

Test proposed lexical items against natural language examples.

12. Conclusions

If a digitally-based RC is to be created, it may need to be founded upon the ideas outlined here and the principles enunciated in Maun (2013). Nothing is, as yet, conclusive. Combining the factors under

consideration into a summary list, we obtain the following:

- The system will be developed for use by a generalist learner.
- Icons, indices, and symbols will be used.
- Semantic primes will be incorporated, including those of Wierzbicka and Goddard, as well identifiable semantic meta-units (affixes).
- The core of a lexicon may well be Basic English, supported by notions drawn from Interglossa and the semantic primes identified above.
- Visual primes will be employed to convey meaning. This will include the dot, straight lines, and curved lines, as well as basic shapes.
- It may be necessary to employ stylistic conventions in the design of all types of character, e.g., human icons being drawn face-on.
- The composition of RC characters may follow formational and reading parameters in the manner of Chinese characters and Maya glyphs.
- Metaphor may be expressed through the use of colour.
- Conceptual support may be available digitally beneath the surface of characters or through animation.
- Syntax will be expressed through the use of the T-bar structure.
- Presentation of messages will be through RSVP on computers and other digital devices.
- Historical examples of Real Character and other visual languages will provide guidance and warnings in the creation of a modern, digital system.

As yet, no fully worked-out, systematic, consistent visual RC has

been created. The systems created by Haag, Bliss, Hanks, Charteris, and Randić offer both good and bad examples of the principles which must govern such a communications system. It is to be hoped that further work will bring such a system into existence.

References

- Charteris, L. 1972. *Paleneo: A Universal Sign Language*. London: Hodder & Stoughton for Interlit A.G.
- Dalgarno, G. (1661) 2001. The Art of Signs. In D. Cram & J. Maat (eds.), *George Dalgarno on Universal Language* 137-289. Oxford: Oxford University Press.
- Demers, R. & A. Farmer. 1986. *A Linguistics Workbook*. Cambridge, MA: MIT Press.
- Dondis, D. 1973. *A Primer of Visual Literacy*. Cambridge, MA: MIT Press.
- Eco, U. 1995. *The Search for the Perfect Language*. Oxford: Blackwell Publishing.
- Esquisabel, O. 2012. Representing and Abstracting. An Analysis of Leibniz's Concept of Symbolic Knowledge. In A. Casanave (ed.), *Symbolic Knowledge from Leibniz to Husserl* 1-49. London: College Publications.
- Everett, D. 2012. *Language. The Cultural Tool*. London: Profile Books.
- Feldman, L. & W. Siok. 1997. The Role of Component Function in Visual Recognition of Chinese Characters. *Journal of Experimental Psychology, Learning, Memory, and Cognition* 23.3, 776-781.
- Frutiger, A. 1989. *Signs and Symbols. Their Design and Meaning*.

London: Studio Editions.

- Goddard, C. 1998. *Semantic Analysis. A Practical Introduction*. Oxford: Oxford University Press.
- _____. 2010. NSM. Available at URL
<<http://www.griffith.edu.au/humanities-languages/school-languages-linguistics/research/natural-semantic-metalanguage-homepage>>.
- Haag, C. 1902. *Versuch einer graphischen Sprache auf logischer Grundlage*. Stuttgart: W. Kohlhammer.
- Hankes, E. 1992. *Eh May Ghee Chah. A Universal Second Language*. Minneapolis, MN: The Hankes Foundation.
- Hogben, L. 1943. *Interglossa*. Harmondsworth: Penguin.
- Houston, S. 1989. *Reading the Past: Maya Glyphs*. Berkeley & Los Angeles, CA: University of California Press & British Museum.
- Jackendoff, R. 2012. *A User's Guide to Thought and Meaning*. Oxford: Oxford University Press.
- Jacobs, H. 1947. *A Planned Auxiliary Language*. London: Dennis Dobson Ltd.
- Large, A. 1985. *The Artificial Language Movement*. Oxford: Blackwell.
- Maun, I. 2013. Towards a 'Real Character' for the Computer Age. *Journal of Universal Language* 14.1, 143-184.
- McCloud, S. 1993. *Understanding Comics*. New York: Paradox Press.
- Munby, J. 1978. *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Ogden, C. 1930. *Basic English*. Cambridge: The Orthological Institute.
- _____. (ed.) 1940. *The General Basic English Dictionary*. London: Evans Brothers Ltd.
- Okrent, A. 2009. *In the Land of Invented Languages*. New York:

- Spiegel & Grau.
- Pankhurst, E. 1927. *Delphos: The Future of International Language*. London: Kegan Paul, Trench, Trubner & Co. Ltd.
- Pei, M. 1958. *One Language for the World*. New York: The Devin-Adair Company.
- Pinker, S. 1995. *The Language Instinct. The New Science of Language and Mind*. Harmondsworth: Penguin.
- Quirk, R. et al. 1973. *A Grammar of Contemporary English*. Harlow: Longman.
- Randič, M. 2010. *Nobel Universal Graphical Language*. Bloomington, IN: XLibris.
- Sapir, E. (1921) 1970. *Language*. Reprint. London: Rupert Hart-Davis.
- Scurfield, E. & S. Liany. 2003. *Teach Yourself Beginner's Chinese Script*. London: Hodder Headline.
- Swadesh, M. 1950. Salish Internal Relationships. *International Journal of American Linguistics* 16, 157-167.
- _____. 1971. *The Origin and Diversification of Language*. Edited post mortem by Joel Sherzer. Chicago, IL: Aldine.
- Ting-Ju, L. & M. Biggs. 2006. A Preliminary Study of Learnable Pictogram Languages. Paper presented at International Conference of the Design Research Society. Lisbon.
- Tseng, W-T. & N. Schmitt. 2008. Towards a Model of Motivated Vocabulary Learning: A Structural Equation Modeling [sic] Approach. *Language Learning* 58.2, 357-400.
- Urdang, L. 1982. *Suffixes and Other Word-Final Elements of English*. Detroit, MI: Gale Research Company.
- _____. 1984. *Prefixes and Other Word-Initial Elements of English*. Detroit, MI: Gale Research Company.
- Wierzbicka, A. 1996. *Semantics: Primes and Universals*. Oxford:

Oxford University Press.

Wilkins, J. (1668) 1968. *An Essay towards a Real Character and a Philosophical Language*. Reprint. Menston: The Scholar Press.