

Fiat Lingua

Title: Designing an Artificial Language: Morphology

Author: Rick Morneau

MS Date: 07-16-1994

FL Date: 12-01-2018

FL Number: FL-000057-00

Citation: Morneau, Rick. 1994. "Designing an Artificial Language: Morphology." FL-000057-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 December 2018.

Copyright: © 1994 Rick Morneau. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Designing an Artificial Language:

Morphology

by [Rick Morneau](#)

September, 1991
Revised July 16, 1994

Copyright © 1991, 1994 by Richard A. Morneau,
all rights reserved.

[The following essay was originally published in the September 1991 issue of the *Linguica APA* (Issue #9). I have made a few minor changes since then.]

In this essay, I will discuss ways in which phonemes can be combined into morphemes (minimal units of meaning), and how morphemes can be combined into words. I will discuss morphology only in a very restricted sense; i.e., the *shapes* of words. I will not discuss inflectional morphology at all. And I will postpone the discussion of derivational morphology to my monograph on [Lexical Semantics](#). As a result, this essay will be somewhat abstract.

Since the morphological rules of a language state how phonemes can be linked together to form morphemes, the morphology of a language will have a strong effect on how easy or difficult it is to pronounce. Fortunately or unfortunately, most people have difficulty with complex consonant clusters, and so words such as *mksjzptlk* are not likely to be part of any language's lexicon (unless, of course, you're from the fifth dimension :-). Even clusters that some people consider simple can be quite a challenge to others. For example, most Indo-European languages allow consonant clusters within a single syllable. English examples of this are the "str" in "string", the "bl" in "blue", the "spl" in "splash", the "sk" in "skip", and the "pr" in "prune". Native speakers of most Indo-European languages have few if any problems producing these sounds, but others who study English find them quite difficult and many never master them. Keep this in mind when designing your artificial language (henceforth AL) if you want your language to appeal to as many people as possible.

A word can consist of one or more syllables. For the purpose of this discussion, a syllable is a vowel or diphthong optionally preceded by one or more consecutive consonants, and optionally followed by one or more consecutive consonants. Thus, for the vast majority of languages, a syllable has the form:

$\{C\}V\{V\}\{C\}$ where $\{\}$ indicates zero or more of the
 enclosed item
 C indicates a consonant
 V indicates a vowel or semivowel

However, very few languages take full advantage of the capabilities of the human vocal tract. In fact, a large majority of the world's languages manage to get by with a subset of the above structure which looks more like this:

[C][S]V[V][S][N] where [] indicates that the enclosed
item is optional
C indicates a consonant
S indicates a semivowel
V indicates a vowel
N indicates a nasal

Thus, the simpler structure will allow syllables pronounced like English "him", "queen", "boa" and "toy", but it will not allow syllables like "hit", "string", "plank" or "flirt". The more complex structure will allow either. Note that the lack of consonant clusters and the requirement that the final consonant be a nasal greatly reduces the number of possible syllables that one can create from a fixed phonemic inventory. However, when two such syllables are juxtaposed, the result is very easy to pronounce. For example, speakers of Indo-European languages can pronounce /gwikto/ as easily as /gwinto/, but speakers of most other languages will find /gwikto/ so difficult that they will often slip in a vowel between the /k/ and the /t/. The nasal /n/ is not a problem because nasals are highly vocalic in nature, and co-articulate very smoothly with the preceding vowel.

If you feel that the second structure is too limiting, you may want to consider a compromise which will be easy to pronounce for most but not all people, and looks like this:

[C₁][S]V[V][S][C₂] where [] indicates that the enclosed
item is optional
C₁ indicates any consonant
S indicates a semivowel
V indicates a vowel
C₂ indicates a continuant
consonant or a nasal

Continuant consonants are fricatives and liquids; i.e., just about everything except nasals, stops and affricates. However, a potential problem shows up here when C₂ of a syllable equals C₁ of the following syllable, as in /bassun/. One solution is simply to insist that the double consonant be audibly lengthened. A second approach is to use only non-continuant and non-nasals for C₁.

Once you've decided on the general shape of a syllable, the next step is to decide how to hook them together to form morphemes and words. At this point, you have two choices: an ad hoc approach or a formal approach. If you plan to borrow morphemes directly from existing languages, then you're limited to the ad hoc approach. Basically, you'll choose your morphemes from existing languages and combine the roots, prefixes, suffixes and infixes to create a word. Esperanto and most of the ALs based on European languages fall into this category.

In a more formal approach, the shape of a morpheme will indicate the role it plays in a word. Thus, a prefix will have a different shape than a root, which will have a different shape than a suffix, and so forth. In fact, if you play your cards right, you will not only be able to split a word into its component morphemes on sight, but you'll also know where word boundaries are, even if there are no spaces or pauses between them. You might say that your morphemes and words are auto-isolating or self-segregating. This, of course, would be ideal if you want to speak to a computer, since you won't have to put pauses between words. [By the way, the problem of isolating words in continuous speech is one of the most difficult that the speech-processing community is now facing. I don't expect a solution any time soon.]

So, how do we create a self-segregating morphology? We do it by insuring that each type of morpheme can always be identified by its shape, and by insuring that each type can occupy only one position in a word. Consider a simple example of an easy-to-pronounce language with only three morpheme types:

C = b, p, d, t, g, k, z, s, v, f
 V = a, e, i, o, u
 S = y, w
 N = m, n

prefix = CSV
 root = CVN
 suffix = CV

word = {prefix} {root} suffix

Thus, examples of complete words would be: *za, ke, tembo, sandu, kwabe, pyobendi, kyusintemda, byupwetu*, and so on. Note that if we removed all spaces and squished them all together, we could easily and unambiguously split them apart. This example, however, has at least one serious flaw. Since the root form is CVN, the maximum number of roots we can form with our phoneme inventory is only $10 \times 5 \times 2 = 100$. Since we'll need much more than that, let's add disyllabic and trisyllabic root forms:

C = b, p, d, t, g, k, z, s, v, f
 V = a, e, i, o, u
 S = y, w
 N = m, n
 SPECIAL = q (English "ch" in "church")
 x (English "sh" in "shop")

prefix = CSV
 root = CVN or CV[N]qV[N] or CV[N]xV[N]CV[N]
 suffix = CV

word = {prefix} {root} suffix

Note that "q" and "x" simply indicate that the root continues with one or two more syllables, respectively. Examples of two-syllable roots would be *binqan*, *temqu* and *saqem*. Examples of three-syllable roots would be *kuxiba*, *tixendi*, *zomxate* and *panxotun*. Next, add prefixes and suffixes and you would have something like *kwabinqandu*, *temqusa*, *pyosaqembe*, *kuxibato*, *fyotixendika*, *zomxatebi*, and *panxotunki*. (With this type of morphology, no one is going to accuse you of being Eurocentric. :-) Note that, even with this small phonemic inventory, you can create 2,250 unique disyllabic roots and 337,500 unique trisyllabic roots.

The above is just one of many possible examples of what can be done with a formally designed morphology. There are many other things that you can do. You can add new forms such as CVC, CV[S]N, C[S]VN, C[S]V[S]N, CV'V, CV'VN, C[S]V'VN, etc. (where the apostrophe indicates a glottal stop)), or you can dedicate specific phonemes for specific purposes as we did above with "q" and "x". Your choices are limited only by the requirements you set for yourself.

[Addendum: An idea that occurred to me after I wrote the above piece was to dedicate a vowel, such as /a/, for exclusive use in creating polysyllabic morphemes. This phoneme would not be used for anything else. For example, a morpheme of type CVN, could be *tun*, *batun*, *kwatun*, *dambyamatun*, and so on. In other words, whenever /a/ appears, it indicates that the morpheme continues to the right. Only the last syllable is used to determine the morpheme's type. See also my (very long!) monograph [Lexical Semantics](#) for yet another way to implement a self-segregating morphology.]

End of Essay