

# *Fiat Lingua*

Title: Three Lesser-Known Tools for Lexicon-Building in Your Conlang

Author: John Quijada

MS Date: 03-28-2017

FL Date: 05-01-2017

FL Number: FL-000044-00

Citation: Quijada, John. 2017. "Three Lesser-Known Tools for Lexicon-Building in Your Conlang." FL-000044-00, *Fiat Lingua*, <<http://fiatlingua.org>>. Web. 01 May 2017.

Copyright: © 2017 John Quijada. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.



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

## Three Lesser-Known Tools for Lexicon-Building in Your Conlang

While there are some conlangers whose favorite part of language-construction is developing the lexicon for their language, many others see lexicon-building as a tedious necessity. Consequently, the process of lexicon-building may sometimes be carried out less meticulously than other components of one's conlang.

Nevertheless, the process of lexicon-building can become more interesting, even fun, if one is aware of the various methods available for word-formation in a language beyond the obvious "thinking up" a lexeme in a one-to-one correspondence with a word from real-world languages. Of particular interest are those methods associated with diachronic/historical contexts. For the conlanger who has labored to design a historical/diachronic context for their conlang's development over time (perhaps involving archaic versions of the language, a parent language, a family of related languages, or geographically-nearby languages as sources of word-borrowing), several common processes of word-formation are available. These methods will be familiar to anyone who has taken a historical linguistics course or has more than a superficial knowledge of etymological principles. For example:

**Direct borrowings (loanwords):** where a word or phrase is taken directly from a foreign language with its meaning (and often its spelling) retained, the only accommodation being that the pronunciation is modified to fit the borrowing language's phonological constraints. English examples include *tête-à-tête*, *je ne sais quoi*, *zeitgeist*, and *schadenfreude*. A more complex example is our word *decal*, shortened from the original *decalcomania*, an anglicized form of the French word *decalcomanie*.

Sometimes, not only the pronunciation, but the orthography of the borrowed word is also normalized, so that the source language is no longer transparent. An English example is *whiskey*, a shortened form of *whiskeybae*, whose original spelling was *usquebaugh*, borrowed from Gaelic *uiscebeatha* 'water of life.'

**Borrowings with semantic shift:** Far more common is the borrowing of words from other languages with a shift in meaning. English is replete with such borrowings, e.g., *muscle* (from Latin *musculus* 'mouse'), *slogan* (from Scots *slogorne* 'battle cry'), *casserole* (a French word meaning 'saucepan'), *futon* (a Japanese word for 'bedclothes/bedding'). English words often shift their meanings when borrowed into other languages, e.g., Spanish *el smoking* 'the smoking jacket', Italian *il camping* 'the campground', French *les waters* 'the toilet, the lavatory'.

**Calques:** Like a loanword, except that the foreign word or phrase is translated morpheme-by-morpheme into the new language. The English word 'skyscraper' has been borrowed as a calque into several other European languages, e.g., French *gratte-ciel* (literally "(it) scrapes-sky"), German *Wokendratzer* ('cloud-scraper'), and Spanish *rascacielos* ("(it) scrapes skies"). As for calques in English, the word 'loanword' itself is an example, borrowed and directly translated from German *Lehnwort*.

**Blendings:** Otherwise known as portmanteau words, formed by morphologically and phonologically merging two words, along with their meanings. Contemporary examples in English include *Frankenfood*, *pixel* ('picture' + 'element'), and *staycation*.

**Conversions:** Also known as functional shift, where a word's grammatical function (part of speech) is changed. For example, English commonly transforms nouns into verbs, e.g., *to accessorize*, *to party*, *to gaslight*.

**Doublets and Triplets:** English has many pairs or even trios of words with subtle (or not-so-subtle) differences in meanings whose origins were derived from the same source at different times. In English,

the source of the first word is often Norman French, while the source of the second word is the same word borrowed later from standard Central French or even Latin. Examples are *cattle*, *chattel*, and *capital*, all derived (the first two respectively via Norman French and Central French) from Latin *capitalis* ‘of the head’) and similarly *captain*, *chief*, and *chef*, each derived at various times either directly or indirectly (through French) from Latin *caput* ‘head’. Another such triplet is *fidelity*, *faithfulness* and *fealty*.

In English, many legal terms are essentially listings of semantic doublets, one term being derived from Anglo-Saxon, the other from Norman French, originally so that the meaning of a legal document would be understood by those both educated and otherwise. Examples include *aid and abet*, *all and sundry*, *deem and consider*, *fit and proper*, *have and to hold*, *terms and conditions*, *son and heir*, *last will and testament*. Examples of triplets include *ordered, adjudged, and decreed*; and *cancel, annul and set aside*.

### Three Lesser-Known But Potentially Fascinating Tools for Word-Formation

In addition to the above-described processes, there are several lesser-known tools for word-formation that many neophyte or even journeyman conlangers who have not studied linguistics formally may be unaware of. Three such tools are the use of **folk-etymology**, **back-formation**, and **phono-semantic matching**. These three processes of word formation can prove to be a fascinating source for building your lexicon. The remainder of this article will respectively examine these three tools of word-formation in natural language, so that you can consider how you might utilize such processes when building your conlang’s lexicon.

#### Folk Etymology

Folk etymology refers to words or short phrases in a language being derived by false etymological assumptions. This may involve a change in either the morphological form of a word and/or its pronunciation, or may simply involve a popular but false belief among a language’s speakers regarding the etymology of a word. The phenomenon is essentially a reflection of ignorance on the part of speakers, driven by a psychological need to alter what are otherwise incomprehensible words, so that they take on a semblance of meaning.

Folk etymology is most often found in regard to foreign borrowings, learned words, old-fashioned/archaic words, scientific names, and place-names. The following are examples, mostly from English:

**female:** from French *femelle* (a diminutive of *femme*), phonologically analogized to *male* by semantic association

**penthouse:** from Middle English *pentis*, in turn from Norman French *pentiz* ‘attached building’ which came from Latin *appendicium* ‘appendage’. The second syllable was analogized to *house*.

**crayfish:** the second syllable phonologically analogized to *fish* from Middle English *crevis*, in turn from Norman French *creveis* ‘crayfish’

**chaise lounge:** from French *chaise longue* ‘long chair’ phonologically (or orthographically?) analogized to *lounge* based on its function.

**hammock:** from Spanish *hamaca*. While the English word shows no folk-etymology, the German form *Hängematte*, Dutch *hangmat*, and Swedish *hängmatta* all literally meaning ‘hang(ing) mat’ as folk-etymologized based on shape and function.

**kitty-corner:** derived from *cater-corner*. The latter word involves an unfamiliar form *cater-*, whereas the former allows for the suggestion of a cat’s furtive movements.

**Step-father, step-sister, etc.:** the prefix is popularly assumed to be the same *step* as in the phrase “one step removed from...”, but in fact goes back to an archaic English word meaning “bereaved.”

**bonfire:** often assumed to refer to a “good fire” from French *bon* ‘good’, but in fact derives from *bone-fire*, referring to the common practice up to the 19<sup>th</sup> century of burning old bones as fuel. Here, the phonological influence of the /nf/ consonant cluster has shortened the long *o* so that the word ‘bone’ has taken on the appearance of French *bon*.

**woodchuck:** from Algonquian *otchek* ‘groundhog’, where the two syllables have been morphed to the closest-sounding English words that bear a seemingly relevant meaning. (Thus giving rise to the popular children’s rhyme “How much wood could a woodchuck chuck if a woodchuck could chuck wood?” Consider folk-etymology as a source for such rhymes and limericks in your own conlang/conculture.)

**bridegroom:** All native English speakers understand the verbal meaning of *groom*, while those familiar with horses also understand the word to refer to a caretaker of horses. So the folk etymology here would entail a man who either grooms the bride or provides care for her horses. The historical derivation is more complex: the Anglo-Saxon form was *brydguma* (from *bryd* ‘bride’ + *guma* ‘man’) which became Middle English *bridgome*. The word *gome* became obsolete by the end of the Middle English period, so that the word came to be popularly changed to *grome* ‘serving lad’, whose meaning narrowed over time to refer to a ‘servant who cares for horses.’

## Back-Formation

Back-formation, also known as juncture loss or juncture metanalysis, refers to the creation of neologisms in a language when the speakers of a language utilize their awareness of its morphological rules to transform an existing word into a previously-unavailable form or part of speech. An example is *pea*, a singular form created from the older English collective plural form *pease*.

Many English verb forms have been created out of misinterpreting the final syllable of a word as being a suffix, when it is actually not. This is common when words ending in *-er*, *-ar*, and *-or* are misinterpreted as having an agentive suffix. Examples:

**burgle:** derived from non-agentive *burglar*.

**peddle:** derived from non-agentive *peddler*.

**lech:** derived from non-agentive *lecher*.

**escalate:** derived from non-agentive *escalator*.

**sculpt:** derived from non-agentive *sculptor*.

The words *swindle*, *edit*, *hawk*, *orate*, and *sculpt* are similarly derived.

Similar to *burgle* and the other forms above, English has many newly-minted words created when speakers interpret a word as containing a “root” that doesn’t actually exist:

***diagnose***: a verb form created from the word *diagnosis*.

***surveil***: a verb form created from the word *surveillance*.

***diagnose***: a verb form created from the word *diagnosis*.

***peeve***: derived from *peevish*.

Other words similarly derived by stripping away what are perceived as affixes are *afflict*, *laze*, *liaise*, *televize*, *revise*, *donate*, *lase*, and *jell*.

What is notable is that these neologisms often fill a void in the language, where speakers sense a seeming lack of a desired form, usually a verb form, and use their innate knowledge of their language’s morphology to fill that void.

Note also that many back-formations never gain long-term legitimacy. So, while forms such as *elocute*, *enthuse*, *evolute*, *aggress*, *attrit*, *evanesce*, and *frivol* are attested in various writings, they have yet to enter English as truly acceptable forms.

And before the reader assumes that any Latinate noun ending in *-tion* or *-sion* is fair game for back-formation, it should be noted that many such words already come with historically-supplied verb forms traceable back to Latin or French, e.g., *administer*, *delimit*, *interpret*, *register*, *revolt*.

### **Misinterpretation of morpheme boundaries in foreign borrowings**

Another (and perhaps more interesting) form of back-formation occurs when speakers create neologisms based on misinterpreted morpheme boundaries. This is especially common with foreign borrowings. For example:

In English, words like *veggieburger* have been derived from *hamburger* based on the false assumption that the latter word described a “burger” made from ham, when, in fact, the word comes from German *Hamburg* + *er*, where the primary morpheme is a geographical reference.

Our words *apron* and *umpire* were originally Middle English *napron* and *noumpere*, where the initial *n* was commonly heard as being part of a preceding indefinite article *a(n)*. The reverse of this process occurred with Middle English *an eute*, now Modern English ‘a newt’.

The Persian word for the game of chess, *shatranj*, is an entertaining example. To Persian speakers this is “a hundred worries” (a fitting name for chess!) based on *shat* ‘hundred’ + *ranj* ‘worry’, derived from Sanskrit *chaturanga* ‘chess’.

Some loanwords in Bantu languages have been misinterpreted by speakers as beginning with a Bantu class/number prefix. An example is the Swahili word *kitabu* ‘book’ borrowed from Arabic *kitābun*.

Because the initial syllable of the word *ki-* corresponds to a common Swahili singular noun-class marker, the root of the word is seen as *-tabu*, thus the plural takes the standard plural marker *vi-* for that noun-class, giving *vitabu* ‘books’.

Similarly, I recall once reading that the English word ‘bartender’ was borrowed into a certain Bantu language (I don’t recall which one) as a plural form *batenda* ‘bartenders’, given that the prefix *ba-* is a plural marker, thus giving rise to the corresponding singular form *matenda* ‘bartender’.

A different error is seen with Arabic loanwords in European languages, where the definite article *al-* ‘the’ is assumed to be part of the noun itself, e.g., English *alcove*, *algebra*, *alchemy*, *albacore*, *albatross*, *alfalfa*, *alcohol*, Spanish *alcalde*, etc.

### Phono-Semantic Matching

Phono-semantic matching refers to phonetically camouflaging foreign borrowings to look like native words. This phenomenon is popular in languages whose speakers are wary of the encroachment of foreign words, yet need new words to express new ideas and phenomena introduced via cross-cultural contact. The following are examples from Icelandic, a language whose construction of phono-semantic matches are actually overseen by a government agency:

*páfagaukur* ‘parrot’, derived from *páfa* ‘pope’ + *gaukur* ‘cuckoo’ to camouflage the Danish source word *papegøje* ‘parrot’

*eyðni* ‘AIDS’ derived from *eyða* ‘destroy’ + *-ni* [nominalizer]

*brokkál* ‘broccoli’, derived from *brok* ‘cotton grass’ + *kál* ‘a plant from the genus Brassica’ to camouflage the English (and ultimately the Italian) source word

*tækni* ‘technology; technique’, derived from *tæki* ‘tool’ + *-ni* [nominalizer] to camouflage the Danish source word *teknik*

An example from Mandarin Chinese is *léidá* ‘radar’ (literally: ‘thunder’ + ‘reach’), while an example from English is the previously mentioned *woodchuck* (<Algonquian *otchek* ‘hedghehog’).

The Arabic language gets into the act as well, particularly when creating words for technological and scientific phenomena, an example being *mutaqaddirah* ‘mitochondria’ from an Arabic root meaning ‘power’, alluding to a mitochondrion’s function as the “powerhouse” within cells. An older Arabic example is *arḍī-shūkī* ‘artichoke’ the two halves of the word respectively meaning “earthly/from the ground” plus “thorny/prickly”.

### Your Conlang’s Lexicon As a Reflection of Your Con-Culture and Con-History

As can be seen, folk-etymology, back-formation, and phono-semantic matching can be employed not simply as linguistic tools for lexicon-building, but also as means for reflecting the linguistic and cultural mindset of your conlang’s speakers in both a historical and political context. This can potentially add greater depth and verisimilitude to your created worlds. Or perhaps conversely, they may inspire you to create a diachronic/historical context for your conlang, or a historical and political context for your con-culture, so as to justify the use of such tools. Either way, have fun with your lexicon!