A World of Words

By Danny Hieber

Words are wonderful things. Despite their ubiquity, their elegance often escapes our notice until we study another language. Then we find words like *s'wcust'aq'cis* from the Atsugewi language, once spoken in Northern California but whose last speaker died in 1988 (though some living members of the community still know a few words and are actively trying to revive the language). This is a motion word, used to describe movement. Motion verbs in Atsugewi have several parts:

s'w- The first piece is a person marker, that tells the hearer who did the acting and what was being acted upon. *s'w-* indicates the first person, "I", but is only be used for factual statements (as opposed to conditional or subjunctive) when "I" am performing an action on something else

cu- In Atsugewi, you can't talk about motion without saying what *caused* the motion, so every motion verb has a cause prefix. *cu-* refers to a linear object (like a stick) moving axially (e.g. poking something), but there are also *ca-*, for movement caused by the wind, or *uh-* for gravity (like falling).

-st'aq- For every movement verb, you also have to talk about the background that the object moves against or into. Here *-st'aq-* is used for runny icky material (if something fell in a pile of manure, or you stepped on a piece of chewed gum), but another example is *-swal-* for linear objects suspended by one end (a shirt on a line, hanging sausages).

-cis Finally, you use a suffix which indicates the direction or location, such as *-cis* (into fire), *-ict* (into liquid), or *-ik* (on the ground).

Put it all together, and you get something like "I caused runny icky material to move into the fire by acting on it with a linear object moving axially" – similar to what would happen, for example, if someone were to prod animal guts into a fire with a stick. How wonderfully evocative![1]

Another great word is *hivomai*, from the Lolovoli language spoken in Vanuatu. In island regions like Vanuatu, directions like north and east are less useful than seaward, uphill, or across the slope, since you can always situate yourself according to these directions on the island. Lolovoli therefore has different verbs depending on whether a person is moving levelly/uphill/downhill, and whether the movement is towards the speaker, addressee, or away from the object spoken about. *Hivomai*, then, is used when something is moving downhill/seaward and coming towards the speaker, like if I were standing on the beach and telling you to come towards me.

Unfamiliar examples like these make one wonder: Just how complex does *my* language seem to others? How difficult is it to understand a word like *institutionalization*, which literally means the act of causing something to become like a thing that has been instituted? In retrospect, the fact that we use such words with regularity—and so effortlessly!—is remarkable.

It becomes even more remarkable when you consider that we frequently don't even know what counts as a word in the first place! For example, is the compound consisting of the two roots

health and *care* one word or two? Competent English speakers seem to disagree. Here is a graph showing the use of *health care* versus *healthcare* in books since 1970:[2]
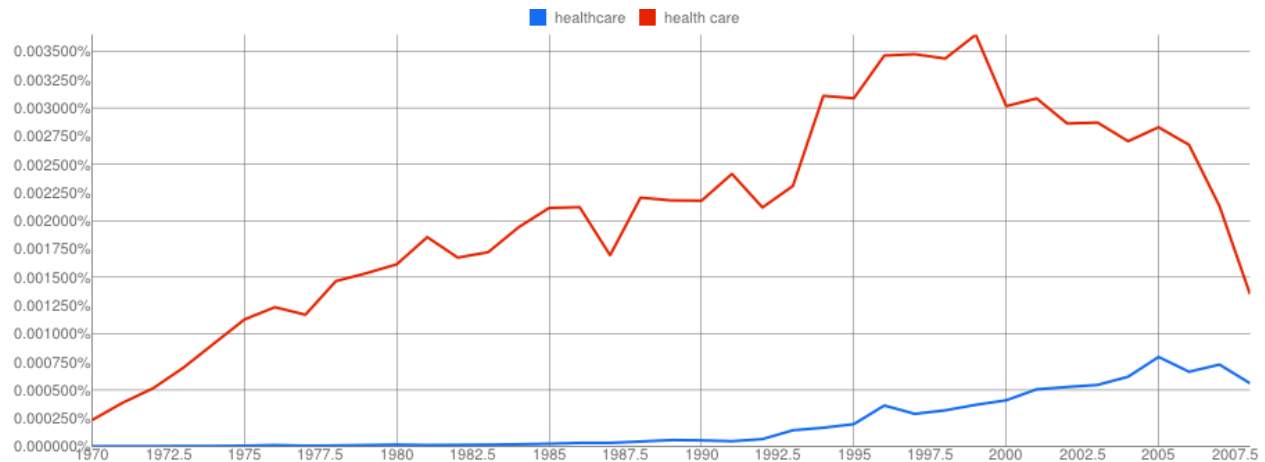
While historically *health care* was considered two words, today you wouldn't be thought insane if you insisted it was one. How do we decide? How do we unlock the arcane secrets of wordhood?

It depends on the habits of speakers. Words do not magically change, come into being, or amalgamate overnight; instead, a few language innovators make the initial leap, and eventually more and more people slowly (or sometimes rapidly) accept the new convention until it becomes the norm. What counts as a word to one person, or in one language, may be different from what counts as a word to other people or languages. There are, however, clues that help guide our assessment—minutiae so subtle that we speakers aren't even consciously aware of them, even though they are rules that every speaker of the language knows.

One of these clues to wordhood is pronunciation. Certain pronunciation rules only apply at the boundaries between parts of words (*morphemes*), and not between whole words. For example, in Chitimacha, a Louisiana language which went dormant in 1940 but is now being learned from archival materials by children on the reservation again, roots sometimes drop vowels when combined with other suffixes:

husa	→	*huskamiig*
five		fifth

Here the /a/ drops off the end of *husa*. But in other phrases this doesn't happen:

husa	→	*husa*	*ne*	*qupa*
five		five	and	two

You couldn't say, for example, *husne qupa* and drop the /a/ like before. Since the /a/ drops off *husa* in *huskamiig* but not *husa ne qupa*, this is good evidence that *huskamiig* is one word while *husa ne* is two.

Hieber, Daniel W. 2013. A world of words. Blog post written for *Linguavore: RVoice* (Rosetta Stone blog). Left the company before it was published.

In English, one clue to wordhood is stress. Consider the difference between a *black bírd* (a bird that is black) and a *bláckbird* (a type of bird). The stress indicates whether these two roots, when combined, should be thought of as one word or two. All speakers of English know this, even if we don't know we know it.

Another criterion often applied to words is that they can stand in isolation. So native English speakers know that *re-* is not a word, but *write* and *rewrite* are. Syntax, which is the order of elements in an utterance, tells us lots about words as well. Many times, the elements in an utterance must occur in a strict order, or always occur together. Verbs in Swahili, a language spoken in East Africa, are good examples of this:

nitakupenda	"I will love you"
ni	I
ta	future tense
ku	you
pend	love
a	(indicates the verb is a statement, and not a subjunctive)

Here, you could not say *nikupenda* and drop the {ta}, because Swahili verbs require a tense marker in most of their forms. Nor could you say *ku-ni-ta-pend-a* and move the *ku* to the beginning of the word. These types of restrictions on the order and presence of elements in an utterance can be good evidence for wordhood.

Even with these guidelines, it's not always easy to tell what's what. For example, is the *-'s* possessive marker in English a word or just a suffix? We know that *-'s* can't stand alone in isolation, so by that criterion it cannot be a word. But it can't just be a suffix either, because suffixes are things you only attach to words. The *-'s*, however, attaches to entire phrases!

[Billy]'s dog
[the man]'s dog
[the man and the woman]'s dog
[the man I saw at the store]'s dog

This *-'s* is actually somewhere between a word and a suffix, and is what linguists call a *clitic*.

Languages also differ drastically in what they allow to be words. Some languages are what linguists refer to as *isolating* languages, where (in the perfect case) every unit of meaning gets its own word, and the words are invariable. Vietnamese is a great example of this:

chị	ấy	đã	quên
she	REFER	PAST	forget

'she has forgotten'

West Greenlandic (an Eskimo-Aleut language), on the other hand, is a polysynthetic language, allowing for words of incredible length:

Hieber, Daniel W. 2013. A world of words. Blog post written for *Linguavore: RVoice* (Rosetta Stone blog). Left the company before it was published.

anigu-ga-ssa-a-junna-a-ngajal-luinnar-simassa-galuar-put
avoid-PASSIVE-PARTICLE-FUTURE-be-no.longer-almost-really-must-however-3PL.STATEMENT
"They must really almost have become unavoidable but . . ."

All this just goes to show the amazing variety of ways that words can be extremely far from what we might expect. To me, this is one of the most thrilling parts of learning a new language. When we encounter exotic new words, like little treasures buried in the sand, we come to appreciate some of the wonderful complexity and diversity that language has to offer.

---

[1] For a full explanation of motion verbs in Atsugewi, check out page 63 of Nicholas Evans' fantastic book, *Dying Words: Endangered Languages and What They Have to Tell Us* (Wiley-Blackwell, 2009).

[2] Data courtesy of Google's Ngram Viewer.