Conlanging 101



by Sai Emrys (ccc@saizai.com¹) with Alex Fink & David Peterson²

> conlanging (n): the art and craft of making your very own language

What?

A constructed language (conlang) is meant to function just like any natural language (natlang)—a complex system for communicating between humans (or, perhaps, aliens or fictional beings).

This is not simply devising a *code*, like Pig Latin, where you take an existing language and superficially change the vocabulary, or a jargon, like hacker English or legalese, where you create new words for a specific topic. Nor is it devising a highly constrained formal language, such as programming languages, which don't need the flexibility to be able to say "I had an awesome time at 26C3".

Conlanging is to linguistics what painting is to art history, or hacking to computer science. It's a way of directly playing with language sometimes just for fun, and sometimes to test out a new theory about how language works with the mind.

Why?

As with all hacking, the motivations vary considerably.

Most conlangers are in it simply for a sense of aesthetic fulfillment. To them³—the artlangers -language creation is an artistic craft, somewhat like model railroad building, costume design, or modern architectural design.

In *Teonaht*⁴, for example, a nine-year old Sarah Higley (aka Sally Caves) was inspired to come

Comments & links greatly appreciated. GPG D6D408A9. Many thanks to Jim Henry, Donald Boozer, Schuyler Duveen,

up with a 'what if' language of flying kittens (the *feleonim*). This humble start blossomed into a lifelong passion with invented languages and the urge to play with language concepts until, today, Prof. Sarah Higley's creation is one of the most-respected examples of modern artlangs.

The Book of Yrlo, a Teonaht cultural text, begins: Keyst, helepmivarn! ta nikkyam perim uom ihhai rrõhhõnt, uom ihhai ferrefib, uom ihhai oy preib ven elepmibjo. ("Consider: there are books that are secret, books that are lost, and books that are known and well-read.")

J.R.R. Tolkien's Quenya⁵ is one of the languages spoken by the elves in his fictional land of Middle Earth. "Nobody believes me when I say that my long book [Lord of the *Rings*] is an attempt to create a world in which a form of language agreeable to my personal aesthetic might seem real," Tolkien complained⁶. "But it is true."

A Quenya greeting is : (رَبُ صَ اللَّهِ عَلَى اللَّهِ اللَّهِ اللَّهِ اللَّهِ عَلَى اللَّهِ اللَّهُ اللَّالِي الللَّاللَّا اللَّهُ اللَّهُ اللَّهُ اللَّهُ اللَّهُ اللَّا اللل síla lúmenn' omentielvo ("a star shines upon the hour of our meeting"). The poem many Namárië (aka Galadriel's Lament) begins: 31 σὸψί ἐΙὰκο τὰ ἐΙὰκοκον είμο ἐμοῦρο τὰ τὰ τὰ τὰ Ταυτίἒ lantar lassi súrinen, yéni únótimë ve rámar aldaron! ("Ah! like gold fall the leaves in the wind, long years numberless as the wings of trees!")

The Klingon language⁸ (**YYT 46 ?7** Cor tlhIngan Hol), set in the Star Trek universe, was created by Mark Okrand based on the warlike culture and snippets of conversation from Klingons in the TV series. Klingon is thus informed by (and sometimes informs) the larger fiction. The language is harsh and guttural, and combines uncommon but natural linguistic features to create an 'alien' aesthetic.

For example, a Klingon would not ask **F** HASLYTIC nug 'oH vuDlIj'e' ("What is your opinion?"), but rather say **१९६७ ** vuDlIj MASSAM ("Give me your opinion!"). Similarly, a fairly polite greeting, **EFFE** nuqueH, is actu-

John Vertical, John Quijada, Mark Shoulson, Freenode #lojban, AUXLANG-L, Sonja Elen Kisa, Philip Newton, James, Larry Sulky, Gary Shannon, David McCann, Henrik Theiling, V, and Kaleissin for their extensive suggestions, comments, edits, examples, nitpicks, & many other improvements. This paper is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 license. Do cool stuff with it, but give me credit, let me know, and share what you make freely.

http://dedalvs.conlang.org/notes/manifesto.php

http://frontiernet.net/~scaves/teonaht.html

http://folk.uib.no/hnohf The Letters of J.R.R. Tolkien, p. 264

Quenya font is Tengwar Formal: http://tengwarformal.limes.com.pl

http://kli.org Klingon font is pIqaD: http://www.kli.org/tlh/pIqaD.html

By contrast, *auxlangers* seek to create an auxiliary language⁹—a language that can be easily learned by anyone, and serve as a neutral bridge between speakers of different languages. Auxlang creators were once far more common, but these days are a minority.

However, successful auxlangs have by far more speakers than artlangs. Where artlangers are more individualistic, often spending a lifetime tinkering with their languages without caring too much about 'finishing' one and acquiring a large speaker community, a certain amount of promotion is necessary for an auxlang to be successful on its own terms.

Most auxlangs are based primarily on European languages, including the most well known auxlang, *Esperanto*¹⁰, with some 100k-2M speakers—and ~1k *native* speakers, or *denaskuloj*. It was created in 1887 by L.L. Zamenhof in reaction to social tension between Russians, Poles, Germans, and Jews in his native Poland. He felt that uniting everyone with a common, neutral language would help to foster harmony. It has since given rise to numerous spinoff languages, books, conferences, songs, and other works.

For example, *Bonan tagon! Kiel vi fartas?* means "Good day! How are you?". One pangram is *Aĥ! Kaŝiĝu vi hejme! Apenaŭ uzeblas ĉi fieca langtordaĵo.* ("Ah! Hide yourself at home. This filthy tongue twister is almost useable.")

Another auxlang is Jens Wilkinson's *Neo Patwa*¹¹. Compared to Esperanto, Neo Patwa is a more international language. It draws vocabulary from English, Chinese, Hindi, Swahili, Spanish, Arabic, Russian, Indonesian, Korean, and Japanese—rather than just from European languages. It also is different from Esperanto and other major auxlangs in that there are no plural forms and no verb tenses.

For example, consider *Do-pela din-tinta na cidya, ta makan pwason*. ("Two blue birds ate

fish"; literally, "Two-thing sky-color bird, it eat fish.") Here, the words for 'two' and 'sky' are from Hindi, 'thing' is ultimately from English (through the Tok Pisin creole), 'color' is from Spanish, the adjectivizing particle na is from Japanese, the pronoun ta is from Mandarin, 'eat' is from Indonesian, and 'fish' is from French. This wordiness, use of compounds, and reliance on metaphorical extensions is the usual tradeoff of having a very small vocabulary.

Engineered languages, or philosophical languages, are both the rarest and the most radical. An engelanger takes a systemic concept and runs with it—with dramatic effects to the resulting language. This isn't a completely exclusive category; many philosophical languages also have artistic or auxiliary elements.

It's hard to describe a 'typical' engelang since they are so divergent, so instead, some examples:

Ithkuil¹², the creation of John Quijada, attempts to pack the maximum amount of information into the smallest space, and expresses levels of human cognition that are usually unexpressed in natural languages, thus minimizing ambiguity and maximizing the precision of meaning.

(oumpeá äx'ääuktëx), for example, means "On the contrary, I think it may turn out that this rugged mountain range trails off at some point." That's a rather extreme level of concision.¹³

Toki Pona¹⁴ (lit. 'language good/simple'), by Sonja Elen Kisa, expresses all concepts using 123 simple root words in a minimalist dadazen manner. More complex terms are either formed using *ad hoc* compounds—e.g. 'adapt' is *ante pona* ('change good')—or self-contradictory—e.g. 'friend' is *jan pona* ('good person'), so 'bad friend' is nonsense.

Toki Pona is good at expressing simple proverbs; *o weka e nimi namako*, for example, means "omit needless words". The motto of toki pona is *ale li pona*—"everything is good".

⁹ sometimes called international auxiliary language, or IAL

¹⁰ http://en.wikipedia.org/wiki/Esperanto

¹¹ http://patwa.pbworks.com

¹² http://ithkuil.net

¹³ Of course, Ithkuil is far more concise to read than to write; even for Quijada, composing a sentence can take hours.

¹⁴ http://tokipona.org

*Lojban*¹⁵ is a *logical language*; like most *loglangs*, its aim is maximum precision and unambiguity. For example, the English phrase "a pretty little girls' school" has a large variety of meanings, indistinguishable without awkward rephrasing—from "a school for girls who are pretty and little" to "a somewhat small school for girls". Each has its own translation in Lojban—e.g. cmalu melbi nixli ckule means "a school for girls who are pretty because they are small", and cmalu je melbi ke nixli ckule means "a small and pretty school for girls".

Lojban also has the flexibility to be vague; for example, mi rinsa lo se vitke means "I/we (will) greet(ed) the/a host(ess)(es/s)"—all information is strictly optional, to be understood from context if left unspecified. The simplest Lojban sentence possible consists only of the word *co'e*, which is a sort of pure metasyntactic variable. In this case, in addition to not specifying any arguments, even the relationship is elided, leaving vague what is happening, never mind when or to whom.

Whence?

Broadly speaking, conlangs arise in two (and a half) different ways.

A posteriori languages are based on an existing language or languages.

Sometimes, they are set in an alternative history; Brithenig¹⁶, for example, is the result of Old Celtic being replaced by Latin, but undergoing the same changes as affected Welsh in the real world. Sometimes, it's as part of a whole family of conlangs (like Tolkein's Elvish languages Quenya, Sindarin, Telerin, Avarin, Silvarin, etc.). Each is interrelated with the others just like natural languages are on Earth (e.g. French, Spanish, Italian, etc. all deriving from Vulgar Latin).

And sometimes it's because taking something and remixing it is just easier than making something entirely new, and the conlanger wants to concentrate on only what they find most interesting. For example, Steven Travis' Tapissary¹⁷ mostly uses English grammar, but has a French-creole sound system and a highly creative writing system.

A priori languages are made from scratch. This is a challenging task—not least because it requires a significant understanding of one's own native language to avoid unwittingly producing something only superficially different from it, with different words but identical structure.

Like any custom hack, however, with a bit of awareness of how the system functions one can make extensive changes to a language to suit one's desires. Perhaps a novel sound system? Etymologies derived from an ontology? A syntax which prevents ambiguous sentences¹⁸? Most engineered languages are a priori for this reason, to avoid being overly constrained.

Some a priori artlangs are associated with fictive cultures of speakers, such as *Kēlen*¹⁹, *Klin*gon, and Taruven²⁰. Other personal languages have no fictional history to go with them, such as *Vabungula*²¹, *gjâ-zym-byn*²², and *Deini*²³.

Of course, making something truly new is hard. So often does someone erroneously announce that they've done so that our community has a standard response: *ANADEW* (A Natlang's Already Done it, Except Worse). No difference between verbs and nouns? No verb 'to be'? Inflection based on where one is standing with respect to the mountain? A single category of words that includes women, fire, and dangerous things²⁴? ANADEW!

Finally, there are natural languages that nevertheless have had a significant amount of intentional human input. Some, like Cherokee and Korean, had a writing system created by a single influential leader and then promulgated to the masses. Other reconstructed languages, like Modern Hebrew, were dead languages for a long time until a linguist sat down and figured out how to use them for modern life... and then were widely adopted for sociopolitical reasons.

And, of course, there are always the prescriptivist²⁵ meddlers. From *l'Académie française* to die

¹⁵ http://lojban.org and http://xkcd.com/191

¹⁶ http://bethisad.com

¹⁷ http://tapissary.com

¹⁸ http://eskimo.com/~ram/lexical_semantics.html

http://terjemar.net/kelen.php

http://taliesin.nvg.org/taruven

²¹

http://billpriceweb.com/vabgram.html http://bellsouthpwp.net/j/i/jimhenry1973/gzb/gzb.htm

http://conlang.dana.nutter.net/index.php/Deini 23

cf. George Lakoff's book by this name

²⁵ Linguists merely describe language as it is really used; telling people how they ought to use language is almost entirely the

Rechtschreibreform, officials do seem to keep trying to 'improve' their native language... or stave off the constant 'degradation' by the next generation. Unlike conlangers, such efforts tend to cling to the old ways of doing things, rather than introduce new features and frequently focus (superficially) spelling, since it's easier to regulate than pronunciation.

Who?

The concept of constructed languages can be traced back as far as the ancient Greeks. Plato's Cratylus dialogue includes an argument on whether words can be arbitrarily assigned, and Athenaeus of Naucratis' work, The Deipnosophists, includes actual snippets of invented words.

The earliest-known working conlang is St. Hildegard of Bingen's 12th century Lingua Ignota, which uses invented words within a Latin grammatical framework. St. Hildegard used her language primarily for devotional purposes. From the sixteenth through the nineteenth century, an increasing number of philosophical languages and auxlangs were invented, along with a handful of sketchy artistic languages used in fiction (e.g. St. Thomas More's *Utopia*, Jonathan Swift's *Gulli*ver's Travels, Edgar Rice Burroughs' John Carter of Mars series).

The first in-depth universal language scheme to be published was Francis Lodwick's A Common Writing (1647). The first conlang to acquire a fluent speaker community was Volapük (1879), an auxlang devised by Fr. Johann Martin Schleyer of Baden; it was fairly popular in western Europe and elsewhere in the 1880s. It was soon superseded by Esperanto (1887), devised by Dr. L.L. Zamenhof of Poland. Esperanto remains the most widely spoken constructed language, although a few of the auxlangs invented since then, especially Ido (1907) and Interlingua (1951), have also acquired a significant number of speakers.

In the twentieth century, especially since the publication of J.R.R. Tolkien's The Lord of the *Rings*, which showcased his Elvish languages Quenya and Sindarin, the proportion of artistic

or fictional languages to philosophical and auxiliary languages has greatly increased.

Conlanging slowly came 'out of the closet'²⁶, especially after the founding of the CON-LANG mailing list in 1991. People who formerly thought that they were the only ones to have such an unusual hobby began to discover fora full of kindred crafters²⁷.

I founded the Language Creation Conference²⁸ (LCC) in 2006, which gives a new platform for conlangers to share their craft.

Of course, conlanging isn't just for enthusiasts. Novels, games, movies, and other creative works often incorporate worlds with their own languages, and increasingly, this work is being done by real conlangers²⁹.

For more on the history of conlanging, see Arika Okrent's excellent book, In the Land of Invented Languages³⁰.

How?

Theory's great, but how does one do such an enormous task? By breaking it up into small ones, of course.

First off: What is the goal of your language? What aesthetic will shape it? What is the culture and world of its speakers? What things do they talk about most? What media do they

A good upfront understanding of the context (real or fictional) in which this language will exist will help to shape all other decisions you make for the language; a language from a matriarchal polyandrous society, for example, will have a very different system for kinship terms than one where women are treated as chattel.

²⁶ Tolkien famously called it 'A Secret Vice' The Monsters and the Critics, pp. 198-223.

^{...} to mixed delight and disappointment, sometimes. As Sally Caves said in her talk at LCC1, "My reaction to CONLANG, when I got on, was: 'This is fabulous! ... I'm not unique any more."

http://video.google.com/videoplay?docid=-3117774526155284922

²⁸ http://conference.conlang.org

²⁹

http://conlang.org/jobs/hire_us.php http://inthelandofinventedlanguages.com http://conlang.org/press.php#ilil

You might choose to have a 'hook'—some interesting limit³¹ or goal to achieve, or a broader sense of personal aesthetic. For example, the speakers of *Dritok*³² have no vocal chords; *Kēlen* has no verbs; and *Láadan*³³ is meant to express a woman's world view.

You also need to decide what scope of work you're aiming for. A naming language is a minimalist conlang that has just enough detail to allow you to create proper names for a fictional setting. For this, you need a sound system (phonology), basic rules for word formation (morphology), and a list of root words³⁴ and affixes with their meanings.

A more extensive language, which allows you to write actual phrases and sentences, requires more attention to the morphology (word-formation rules) and also *syntax* (rules for arranging words within sentences). A fully speakable language, in which you can write or speak at length on arbitrary subjects, will require attention to pragmatics and stylistics (rules for structuring of texts and conversations, differentiation of texts of different genres, etc.), as well as a much larger vocabulary.

Care in designing the *semantics* of your language (the way its words map to parts of reality and imagination, perhaps in ways interestingly different from any natural language³⁵) is good for any kind of conlang, but especially important for artlangs. Real languages have very few words that mean the exact same thing as another language's words—especially when you consider prepositions and idioms. For instance, English body and German Körper both can mean 'a live human body', but in German it also means 'field' (as in math), and in English it also means 'dead body' (vs. *Leiche*). Then try contrasting English to vs. German zu...

To give a sense of how a typical artlang answers these questions, for the rest of this section, we'll be using examples from David Peterson's language Zhyler³⁶, which he began in 2001. His goal was to implement a vowel harmony system similar to Turkish's, while using no adpositions³⁷ whatsoever, instead relying solely on a large system of noun cases (57, compared to the 4-14 most case languages usually have). In addition, Peterson wanted to create a noun class system as extensive as Swahili's, yet entirely unique amongst natural and created languages (for example, there are separate classes for land mammals, other nonmammalian land animals, human beings without titles, and human beings with titles).

Now that you know what to make, where do you start? Generally, from the basic building blocks of language: sounds. It is sounds that are primary, not letters³⁸; every human is well on their way to mastering at least one language before they even learn what writing is, and of course preliterate societies get by fine without writing.

The International Phonetic Alphabet³⁹ (IPA) is the standard system among linguists for transcribing sounds. CONLANG-extended X-SAMPA⁴⁰ (CXS) is used among conlangers for rendering the IPA in plain ASCII.

The primary ways consonants are arranged is by where in the mouth they are pronounced, or the place of articulation (PoA—from the lips to the throat), and the manner of articulation (MoA), like whether the tongue fully *stops* the air, vibrates against the palate (a *fricative*), etc. There are also other factors, such as whether the vocal chords are vibrating (voicing) or the sound is routed through the nose (nasalization). Vowels are similar, but also involve the relative *height* of the tongue in the mouth and whether the lips are rounded.

The sound of a language—its phonaesthetics, or sprachgefühl—contributes a tremendous portion to the perception of a language. As John Quijada said⁴¹ at the 2nd Language Creation Conference, "Phonaesthetics is the reason that Aragorn, having defeated the evil lord of Mordor, becomes King of the West and takes

language-personalities

³¹ http://equilibrium-economicum.net/twokindsoffreedom.htm

http://kryslan.pbworks.com/Dritok http://podcast.conlang.org/2009/02/dritok-the-sound-of-no-

voice-speaking 33 http://laadanlanguage.org

Jeffrey Henning recommends devising about 150 root words for a naming language: http://langmaker.com/ml0102.htm http://nkuitse.com/conlang/glosses

³⁶ http://dedalvs.conlang.org/zhyler

The generic name for prepositions, postpositions, etc.

³⁸ For sign languages, the building blocks are elements of signing, like hand shape and movement. Linguists use the term 'phonology' for both, because they're so similar. Creatures that use something else (e.g. scent?) would probably still have a similar system.

http://web.uvic.ca/ling/resources/ipa/charts/IPAlab/IPAla b.htm

http://theiling.de/ipa

http://www.let.rug.nl/~kleiweg/L04/Tutorial/xsamchart.gif http://podcast.conlang.org/2009/11/lcc2-john-quijada-

the name *Elessar Telcontar*, rather than having defeated the evil lord of *Ailuanyemarë* and being crowned under the name Kratchmurg Brogdoodle."

Languages' use of sounds can be described at two levels: *phonetics*, the actual sounds (phones) that speakers produce, between square brackets (e.g. [ph]42), and phonology, the more abstract underlying level of phonemes, written between slashes (e.g. /p/). Phonemes are the contrastive units of language, and different phonemes can make the difference between different words.

In Zhyler, the *phonetic inventory* (the list of all the sounds that occur in the language) consists of the consonants [p b t d t $\int dz k g f v \theta \delta s$ $z \int z x y m n n n l r j w$ and the vowels [i y w u e ø a o]. This is fairly typical in size and distribution.

Each phoneme can have more than one allophone, or way it can be pronounced. Which allophone you use is usually determined by the surrounding sounds; the difference between the allophones is never meaningful. For example, in English, the /t/ in stop is pronounced without aspiration (a little puff of air), whereas in *top* it is⁴³. Thus we say that in English, $[t^h]$ is an allophone of /t/ that occurs at the beginning of a word.

In Zhyler, there are ten fricative phones. However, four of them—x, y, f, and θ —are not phonemes; that is, they appear only as pronunciation variants of other phonemes.

This is controlled by two rules⁴⁴:

- 1. Spirantization: /k/ and /g/ become [x] and [y] between two vowels.
- 2. Devoicing: $\langle v \rangle$ and $\langle \delta \rangle$ become [f] and $[\theta]$ at the end of a word or next to a voiceless sound.

Thus, /mekel/ 'you are' is pronounced [mexel], and /tiv/ 'swollen' is pronounced [tif]. And a word like /werven/ 'wolf' in Zhyler could never contrast with a word like */werfen/45. Zhyler, like most naturalistic languages, has many more phonological rules affecting the formation of words.

Natural languages exhibit recurring tendencies in their inventories of sounds; they tend to favor symmetric inventories without lots of gaps (e.g. if /t d g/ are there, expect /k/ as well), and if they include complicated sounds (e.g. nasal vowels) they tend also to include more basic versions⁴⁶.

Languages have restrictions on which strings of phonemes are pronounceable, known as phonotactics. What are the possible shapes of syllables? Some languages only allow (C)Vsyllables, i.e. one Vowel (the *nucleus*) with an optional Consonant before it (in the onset) and nothing at all allowed after it (in the *coda*). Other languages allow much longer syllables, with clusters of consonants in both coda and onset, though with restrictions; e.g. English allows /tr-/ but not /tn-/.

Zhyler's syllables are (C)V(C)—that is, it can have up to one consonant in both the onset and the coda (e.g. /bul/ 'thin'; /za/ 'that'). So whereas in English there are words like 'blow', 'start' and even 'strap', a word of Zhyler will never begin with anything more than a single consonant.

Even though sounds are primary, you still do need to write your language down somehow. This might be any or all of the actual orthography used by your language's speakers (if they have one), a romanization invented by the linguists who first study your language, or a romanization simply for your own use.

Romanizations are pragmatic tools, which exist so that you can discuss your language's words and phrases without requiring your readers to learn a whole new script. So in designing a romanization, be systematic, and don't do anything too unique. By contrast, there is great variation in *orthographies*. An orthography is the method by which speakers of a given language write their own language. For example, Zhyler uses⁴⁷ an *alphabet*, like English:

⁴² IPA font is Gentium: http://scripts.sil.org/Gentium

⁴³ To test this, try pronouncing each with your hand in front of

⁴⁴ In more formal style: C[+velar] > [+cont] / V_V and C[+cont, -stri] > [-voice] / _,C[-voice] or _#
45 Linguists use * to mark hypothetical words that are

ungrammatical.

⁴⁶ For lots of examples of inventories of natural languages, see the UPSID database:

http://web.phonetik.uni-frankfurt.de/upsid.html 47 http://dedalvs.com/zhyler/orthography.html

Orth. ·ÄŢURŢW OVAUV OAĀVİM AĢĪĀŢAXŢWPO Rom. Detčer manðat usþak nogenlerüm. IPA [dettʃer manðat usθak nogenlerym] "I took the bowl from the kitchen to the dining room."

But orthography is not limited to simple alphabets (orthographies where one glyph stands for one sound). Some languages (e.g. Japanese) utilize systems whereby each permissible *syllable* has its own glyph. Using an example from Peterson's *Aaalis*, in sact are three glyphs, the first standing for the syllable [ooa], the second [sa], and the third [ja].

Still others utilize glyphs known as *logo-graphs* to illustrate concepts or words, which may be derived from drawings of what they represent, or a combination of independent phonetic forms plus a determinative of some kind, or may be completely abstract symbols. In Peterson's $Kamakawi^{48}$, logographs are used for many words, such £ *leta* 'wing', Я *kala* 'to talk', ħ *nule* 'bridge', and ħ *opeku* 'trouble'.

The orthographies of most natural languages, though, are quite a bit more complex than any label would indicate, when analyzed holistically⁴⁹. A full orthography often includes punctuation, a number system, methods for emphasis, irregular spellings, and non-standard variants (e.g. "i cant go 2 teh stor rite now"). Designing an orthography doesn't end when one has come up with a way to represent each sound found in the language graphically.

Sounds and writing just scratch the surface of language, of course. You need to understand the internal composition of words (*morphology*), and how words fit together to form sentences (*syntax*; together, *morphosyntax*).

Languages tend to differ in how much information they pack into each word. *Isolating* languages have long sentences with short, atomic words (e.g. Chinese and Vietnamese), whereas *synthetic* languages are characterized by very long, complex words, with some words being able to express the content of an entire sentence of English (e.g. Inuktitut and Swahili). Some synthetic languages are *agglutinative*, building their words out of chunks

each expressing one component of the meaning; others are *fusional*, and their words can't be chunked this way (e.g. in Spanish *amé* 'I loved', the *-é* can't be split into a piece meaning 'I' and a piece meaning past tense).

It's up to the language creator to decide how much information to pack into each word, and also how that information will be represented. No natural language, for example, is entirely free of affixation. But a given language may have prefixes (the 're-' in reproduce), suffixes (the '-ing' in running), infixes (the '-ma-' in edumacation) or circumfixes (the 'ge-...-t' in German gesagt 'said'). Many languages use a mix of these; some use only one strategy.

Zhyler is an exclusively suffixing agglutinative language. To create new words, or modify existing ones, one or more suffixes are added. One might start with [gur], the Zhyler root meaning 'strong', and end up with [gurjurkillerrizymvit], which means, "Was I not still very strong?"

This is typically shown with an *interlinear*⁵⁰: gur -jur -kil -ler -riz -ym -vit strong -INTNS -DUR -PST -NEG -1 -Q

In the previous example, you'll notice that there have been quite a few suffixes added to the verb to create a new verb out of the root. In order to translate that one word, six words are needed in English. The difference lies in which *morphological categories* are realized on the verb, and which are relegated to separate expressions. In Zhyler, there are more categories realized on the verb than in English.

Conlangs can be distinctive both by marking *more* (e.g. a separate *dual* form for exactly two things) or *fewer* (e.g. not specifying gender on pronouns) things than familiar natlangs⁵¹. In Zhyler, unlike English, there is no marking for definiteness; for example, *wervener matum* means "I see (a/the) wolf".

⁴⁸ http://dedalvs.conlang.org/kamakawi/orthography.html
49 cf. Trent Pehrson's taxonomy of writing systems:
http://idrani.perastar.com/ISMS_orthography.htm#writings
ystems

⁵⁰ This can be combined with the orthography, romanization, and gloss lines (like in the previous example). The abbreviations and format are standardized according to the Leipzig Glossing Rules: http://www.eva.mpg.de/lingua/resources/glossing-rules.php. The ones used here are intensive, durative, past tense, negative, first person, and question.

⁵¹ The World Atlas of Language Structures contains a huge number of examples of different morphologies: http://wals.info There's also a conlang version: http://cals.conlang.org

Once a language has its words, the creator needs to decide how to string them together. In English, for example, the *subject* (S) of the sentence normally comes before the verb (V), which is followed by the *direct object* (O). In Zhyler, on the other hand, the order of O and V is the opposite (like in Latin). In English you have "The man (S) sees (V) the wolf (O)"; in Zhyler, that would be "Sexa (S) wervener (O) mat (V)."

The ordering of S, O and V is a common way to distinguish between language types (*typology*). Cross-linguistically, SOV is the most common, followed closely by SVO. There are a number of languages with an order of VSO (e.g. Hawaiian), and a modest amount with a VOS word order (e.g. Malagasy), but very few have orders of OSV or OVS (the latter is notable for being the word order of Klingon).

Typically in a natural language, the various elements already mentioned hang together in an important way. For example, if O precedes V in a language, it's likely that adjectives will precede the nouns they modify. The reason is that the verb and the noun are the *heads* of their respective phrases, and heads tend to occur in the same place in every phrase (either first or last). In Zhyler, a head-final language, O precedes V, adjectives precede nouns, relative clauses precede the nouns they modify, and possessors precede possessed nouns. It's also no accident that Zhyler is a suffixing language, with tense elements, cases and noun classes coming after the roots they modify. The result is a language that has internal linguistic consistency, which is precisely what linguists find more often than not in natural languages⁵².

So you have the skeleton of a language⁵³. Now what? Use it!

The usual way one takes a language sketch and 'promotes' it to the status of full conlang is by translating a short text. The most common are the Tower of Babel story from Genesis 11:1-9⁵⁴, the Lord's Prayer, the North Wind and the Sun, and the UN's Declaration of Human Rights. There's certainly no limit,

52 Of course, all natural languages also have exceptions to this kind of common trend; not having exceptions would be just as unnatural as having too many.

53 This text has of course only discussed a small part of what goes into a real language, and omitted many details. If you want to learn more, see the appendix.

54 http://langmaker.com/babel/babel.htm

though. A group of translators from the Klingon Language Institute have translated Shakespeare's *Hamlet* and the epic of Gilgamesh into Klingon⁵⁵, and their current project is the Old Testament of the Bible. Amibitious conlangers also make original works, like LoCoWriMo⁵⁶ and Paul Purgess' *Mna Sipri Cilama*⁵⁷.

One activity unique to the online conlanging community is what's known as the Conlang Relay⁵⁸. In a Relay, the first participant creates a short text in their conlang, and then passes it—with grammar notes—to the next participant, whose job it is to decode the text and translate it into their own conlang. That is sent on to the next participant, and so on, until the text has passed through often more than twenty languages. By the end, the text has usually become something quite different from what it was, and each participant has learned something about how to translate a text, and how to teach others to work with their language

Of course, just because one is able to translate a text in a language, that doesn't mean the language is complete. English has ~300k-1M words, and the average adult knows about 10k-60k words⁵⁹. Creating even 5,000 words is a tall order and can take a lot of time, so realistically, a conlang's lexicon is never complete.

Outside of vocabulary, many conlangers find ways to expand their languages over time. For a language spoken by an imagined group of speakers, new dialects may emerge over time, or perhaps different social registers. Advanced conlangers will create a *proto-language* from which future languages are derived in the way that Spanish, French and Italian ultimately derive from Latin⁶⁰.

Once a language is ready to be presented, the usual method is to make a website⁶¹. Even if you don't have your own webspace, there are conlang wikis which let you document a language in its entirety for free⁶². Reading others'

⁵⁵ http://kli.org/stuff/Hamlet.html

http://kli.org/stuff/ghIlghameS.html http://wiki.frath.net/LoCoWriMo

⁵⁷ http://paulburgess.org/msc.html

⁵⁸ http://dedalvs.com/relay/oldrelays.html

⁵⁹ depending on how exactly you define 'word' (it's hard!)

⁶⁰ This is a major simplification of a fairly complex language family.

http:// /en.wikipedia.org/wiki/Romance_languages#History
 The Ithkuil website is an example of 'gold standard' level documentation.

⁶² http://wiki.frath.net

work will give you a better idea of how to do this well.

Of course, there really is no such thing as a 'finished' language, and the process presented above is normally not done in an exact logical sequence. In real life, it's a huge tangle of revision piled on revision—figuring out which early

vocabulary is now 'incorrect' and needs to be either changed or *retconned*⁶³, tweaking this or that grammatical rule or word root, etc. All good conlangs have gone through many such revisions, so don't feel intimidated that your first try isn't perfect!

Me?

So, what am I into? With my organizational hat off, I'm mostly an engelanger. I have two current primary conlanging interests.

One, developed together with my partner Alex Fink (an artlanger), is a *gripping language*⁶⁴; it has no proper name so far, the medium being unique enough to identify it: it is entirely tactile, consisting of motions made by the two conversants' clasped hands. Its goal is to be an intimate and nearly undetectable language, allowing two people to converse freely and covertly, especially about other people and situations at hand.

The phonology consists of presses on various points on the back and side of the other person's hand using one's fingers and thumb. Such finger presses can be made very lightly, and are easier to feel than to see; even in the unlikely event that someone is staring directly

at the your hands, they're unlikely to discern anything. Several presses can be made simultaneously; there are 125 permitted *chords* of presses, perhaps the analogue of spoken phonemes or syllables. Sequences of chords are strung together into words, with most common words being no longer than two chords. The language has an orthography reminiscent of musical tablature, as well as a romanization

for convenience; the example, transcribed 3a23a 5'45e 35a24a 5'35, means "(they say) a snake perceived a mouse"⁶⁵.

The gripping language is morphologically mostly isolating and syntactically lean, with many structural words omissible where context makes the sense clear. Only

two people can be involved in a gripping conversation, and the channel is inherently asymmetric in that someone's thumb will be on the outside of the clasping; therefore, the pronoun system is built not around first and second person pronouns but 'thumb-outside' and 'thumb-inside' ones. There are also several special series of pronouns to refer specifically to other people present, things they said, etc.

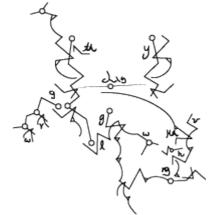
The lexicon is arranged taxonomically, with special emphasis placed on families of words whose meanings smoothly vary on a scale as the thumb moves within its range (e.g. from 'deeply asleep' with the thumb at one end to 'wide awake' at the other). There's also a special mode for encoding English words, used for proper names and one-time borrowings.

My other main interest is in *non-linear writing* systems (NLWSs). A NLWS is to normal text as a graph is to an array; it uses two dimensions to visually encode the relationships between words, and has lots of interconnections between them, rather than merely stringing them out in a line.

Fundamentally, all natural orthographies are really *transcription* systems for the spoken language⁶⁶. They're not optimized for their medi-

um (a 2D writing surface), let alone for the media that are possible with computer-driven interfaces (e.g. reader interaction). A NLWS, by contrast, doesn't care much about spoken language; instead, it primarily tries to be an optimized system for conveying linguistic information in 2D.

In a NLWS, you can do some things you can't in linear writ-



⁶³ From 'retroactive continuity', the practice of fixing something in historical canon by changing the rules

⁶⁴ http://000024.org/conlang/gripping.html

⁵ http://youtube.com/watch?v=7DRnXASa1VM http://bit.ly/gripping_examples

^{66 (}modulo historical cruft and semantic components of logography)

ing systems. You can have the *visual* structure of an argument correspond to its *logical* structure⁶⁷. You can have fractal-like text, where 'zooming in' on a node reveals more detail—e.g. if the top-level story of *Romeo and Juliet* is "boy meets, woos, loses girl" then zooming in on *boy* would give Romeo's personal history. You can have stories (and poetry) that isn't dependent on restricting the reader to reading in a particular order, but rather has its 'punch line' be based on the gestalt of comprehending the overall structure. And of course there's plenty of planar graph theory to explore, for people interested in math.

For me, this is primarily a theoretical and design problem, but others—particularly Schuyler Duveen, with his language *Ouwi*⁶⁸—have actually implemented their versions of this concept, to my delight.

W00t!

Hopefully this has given you a taste for the unique craft of language creation.

If you find yourself interested, I encourage you to just try it for yourself. It's not hard to start (though to become a master certainly takes practice), and the community is very supportive of newbies.

As you explore this, consider joining⁶⁹ the nonprofit Language Creation Society. We do lots of stuff for conlangers: conferences, podcasts, publishing, hosting, jobs, and more.

In any case: happy hacking—or as we say, fiat lingua!

More?

Conlang Mailing List (CONLANG-L) http://listserv.brown.edu/archives/conlang.html Founded in Sept. 1991; oldest & most active conlanging mailing list; many linguists; original usage of the word 'conlang'

Mark Rosenfelder, *Language Construction Kit* (LCK) http://zompist.com/kit.html
Step-by-step guide conlanging for beginners; in English, Portuguese, Italian, and German.

Language Creation Society (LCS) http://conlang.org

Runs Language Creation Conferences, LCS Podcasts (w/interviews of Okrent, Higley, & Payne), etc.

Omniglot

http://omniglot.com

Major online collection of writing systems, both historical and fictional (see Alternative section)

Arika Okrent, In the Land of Invented Languages ISBN 978-0385527880

Excellent history of conlangs: Esperanto, Lojban, Wilkins' philosophical language, Blissymbolics, Klingon, etc.

Thomas E. Payne, *Describing Morphosyntax* ISBN 978-0521588058

Guide to documenting a real language. Excellent intro to the range of grammatical options. AKA 'the conlanging bible'.

Pablo David Flores, *How to Create a Language* http://www.angelfire.com/scifi2/nyh/how__all.html Similar to the LCK, with a different approach. Explains a lot of linguistic concepts.

Sarah Higley, Hildegard of Bingen's Unknown Language ISBN 978-1403976734

Scholarly yet very readable re. Lingua Ignota, its cultural & historical context (w/ comparisons to modern artlangs)

Zompist Bulletin Board (ZBB)

http://spinnoff.com/zbb

PhpBB; younger, more informal and broader than CON-LANG, about as active; run by Rosenfelder

Auxlang Mailing List (AUXLANG-L) http://listserv.brown.edu/archives/auxlang.html Split from CONLANG, exclusively for discussion of auxiliary languages.

That's just the biggest ones! For more, visit the Conlanger's Library ⁷⁰ and the LCS press info page ⁷¹.

⁶⁷ This itself is fairly radical; e.g. it breaks the notion of 'rephrasing'.

⁶⁸ Example is complete 1st chapter of Tao Te Ching http://ouwi.org/writing.html#taoteching

⁶⁹ http://conlang.org/members.php