

A Brief Introduction to Constructed Languages

An essay by Laurier Rochon

Piet Zwart Institute : June 2011

3751 words

Abstract

The aim of this essay will be to provide a general overview of what is considered a "constructed language" (also called *conlang*, *formalized language* or *artificial language*) and explore some similarities, differences and specific properties that set these languages apart from natural languages. This essay is not meant to be an exhaustive repertoire of all existing conlangs, nor should it be used as reference material to explain or dissect them. Rather, my intent is to explore and distill meaning from particular conlangs subjectively chosen for their proximity to my personal research practice based on empirical findings I could infer from their observation and brief use. I will not tackle the task of interpreting the various qualities and discrepancies of conlangs within this short study, as it would surely consist of an endeavour of its own.

It should also be noted that the varying quality of documentation available for conlangs makes it difficult to find either peer-reviewed works or independent writings on these subjects. As a quick example, many artistic languages are conceived and solely used by the author himself/herself. This person is obviously the only one able to make sense of it. This short study will not focus on *artlangs*, but one would understand the challenge in analyzing such a creation: straying away from the beaten path affords an interesting quality to the work, but also renders difficult a precise analytical study of it. In many ways, I have realized that people involved in constructing languages are generally engaging in a fringe activity which typically does not gather much attention - understandably so, given the supremacy of natural languages in our world. The reader should be aware of this while reading this essay.

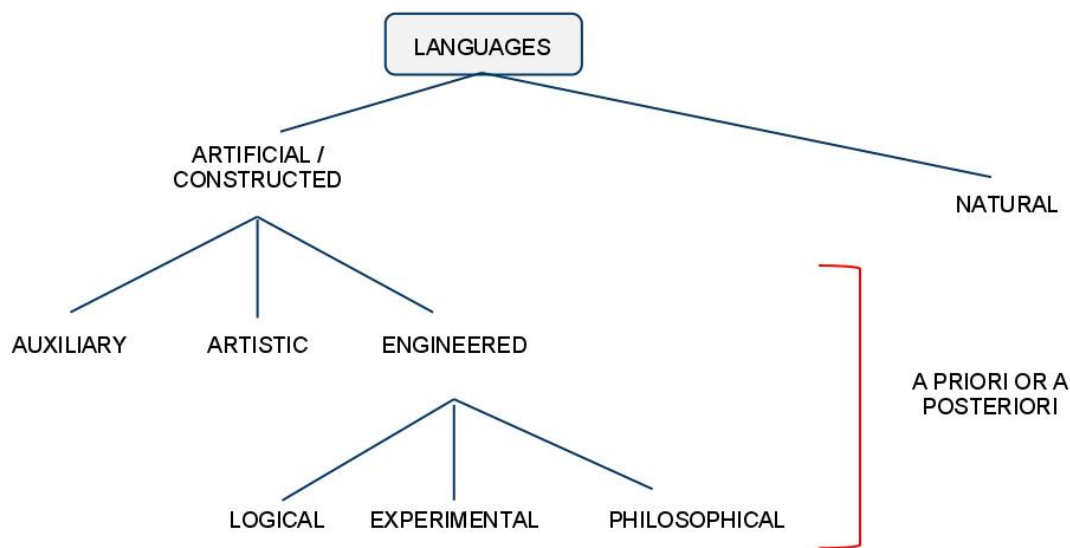
Introduction

Why would anyone want to create a new language, considering the diversity, quantity and complexity of existing ones? Which ideals would one strive for in brushing aside such a long history of collective work in making natural languages evolve? The field of constructed languages is, to a certain extent, built atop of the logical positivist ideas, a group which formed originally in the early 20th century in Vienna. Their inclination towards rationalism and the rejection of metaphysics as being nothing but meaningless (Weinberg, 2001, p.1) has seemingly paved the way for the underlying structures of contemporary constructed languages. They offered to analyse all concepts and propositions through the lens of logical analysis, in a similar fashion that positivism and pragmatism had done previously (Weinberg, 2001, p.2).

I believe this question of *why* one would rewrite the rules of communication from scratch, to be perhaps the most important one to ask. To be in a position to answer it, it is essential, I think, to put oneself in the role of the maker, and try to understand its motivations and reasoning. In this sense, the following text is certainly more a venture into *what* has been done and *how* it has been done, than an answer to

the impending question *why* it was done. Perhaps a thorough breakdown of logical positivism would be beneficiary for the introduction of the following topics, but this would simply prove too lengthy, unfortunately. My goal in presenting this work in an inverse chronological fashion, that is to say the product of a philosophy rather than the philosophy itself, is to demonstrate it using practical examples which feed into my own research experiments, rather than laying out a purely theoretical research paper on the topic.

Constructed languages, natural languages



Constructed languages are languages that have not evolved naturally in use and design. They have been meticulously planned out, or “deliberately designed for a purpose by one person or a small group of people over a relatively short period of time” (Brown & Ogilvi , 2008, p.75). In this sense, languages deemed *natural* are constantly in evolution and update/renew themselves in a typically organic fashion. Their use and application in the real world form the basis for what later becomes accepted as rules. On the other hand, constructed languages usually have specific goals in mind, and vary enormously in their features depending on these goals. They are typically mapped out and rationalized much before their actual form takes shape.

Generally, constructed languages can fit in one of two categories : *a priori* and *a posteriori* languages. The latter have their grammar and vocabulary based on existing languages and the former have their grammar and vocabulary “formed *ad hoc*” and not derived from existing languages (Brown & Ogilvi , 2008, p.77).

For example, Esperanto is *a posteriori* language as it is essentially a mashup of various existing languages while Speedtalk, invented by Robert Heinlein for his novel *Gulf* (1949) is *a priori* since its underlying structure is purely logical, and does not draw at all from existing languages. These two “categories” should perhaps be considered more as “properties”, since they are fairly fluid and part of a larger umbrella of language classifications.

Another general classification that is made in constructed languages concerns their intended goal. Esperanto was created in 1887 and is the most widely used constructed language in the world. [1] It is considered *auxiliary* since its main purpose is to be used by the largest possible number of people, to become widely utilized (it has been recognized by UNESCO in 1954 [2]). Languages such as Interglossa (devised by Lancelot Hogben) and Ido (based on Esperanto) also fall under this category, as they are meant to become “international” languages, or universal second languages. My research has led me to believe that the history of auxiliary languages is quite extensive (echoing Brown and Ogilvie’s thoughts), and therefore I will not venture into their details here. Other categories of languages defined by their goals are *engineered languages*, usually created to experiment in the fields of logic, philosophy or semantics, and finally *artistic* languages, usually intended for personal, aesthetic or mystical reasons [3].

Engineered languages : philosophical, logical, experimental

Furthermore, we can divide engineered languages into a few more subcategories, notably *logical* languages, which are based “on predicate logic rather than one or more common grammatical principles” (Brown & Ogilvi , 2008, p.77) (the most famous probably being *Lojban* by the Logical Language Group Inc.), experimental languages which vary too much in form and aims to explore here (one example could be François Sudre’s *Solresol*, based on a seven-note musical scale (Brown & Ogilvi , 2008, p.76)), and finally philosophical languages, which usually encapsulate some form of political statement reflected in their design, form and/or function. Themaat, in his *Formalized and Artificial languages* offers the following explanation for their formation :

[...] all ideas of the human mind should be analysed into absolutely simple ones. If then symbols were chosen for the simple ideas, the expression for a compound idea was obtained by adding the symbols for the simple ideas, from which that idea was composed. [...] The expressions for the compound ideas would then be obtained by multiplication of the expressions for the simple ideas of which they consisted. Such a language would make the structure and the connection of notions perfectly clear; reasoning would become calculation. (1962, p.320)

He then continues, noting that no philosophical language appears to be usable, “even to the slightest degree”:

[...] one would have to know the structure of the Universe exactly ; [and] even if such a comprehensive world-theory existed, the classification of notions based on it would be so complicated, that the language would be more difficult to learn than the vocabulary of a natural language to be learned from the beginning.” (1962, p.321)

Nonetheless, several interesting attempts at creating usable philosophical languages can be observed, *Láadan* being one example. Created in the mid 1980’s by Suzette Elgin, the language she devised in support of her science-fiction book *Native Tongue* is based on muted group theory and built to counter male-centric languages which impose restrictions on women [4]. Another example is *Toki Pona* created by Sonja Elen Kisa. This language is meant to be as simple as possible (just over 100 root words, 14 phonemes), and focuses on the common cultural denominators rather than trying to figure out the intricate details of bridging linguistic gaps between different speakers[5]. It is heavily influenced by Taoism and highly interpretive, encouraging its speakers to use body language and intonation to make up for the lack of grammatical and syntactical complexity.

At the other far end of the spectrum, we find the logical languages mentioned earlier, which are based on predicate logic and absolute syntactic clarity. For example, *Lojban* claims to be designed to be used by people in communication with each other and possibly computers (in the future), culturally neutral, grammatically unambiguous, to have phonetic spelling and no exception to its rules [6]. Usually, logical languages like *Lojban* do not strive to streamline design aspects to facilitate learning.

A slightly more extreme example of this is *Ithkuil*, a cross between a logical and philosophical language, according to its author, John Quijada [7]. Ithkuil was designed as a highly efficient language, permitting to “blend a high degree of communication of cognitive intent and meaning with a high degree of efficiency, i.e., to allow speakers to say a lot in as few syllables as possible” [7]. This language is certainly a very compressed one, as a very extended phonological field was created for it. This allows for very specific sounds to carry much meaning on their own, which in turn creates the equivalent of English “words” with single syllables. Entire “sentences” are made of a few characters, and complete paragraphs with a single line of glyphs. Admittedly, Ithkuil is incredibly efficient, but also insanely complex to use orally, and to learn in its written form. It would also require ideal circumstances for the people speaking it, as the array of possible sounds is so vast and complex, that having a conversation in a bar, for example, is almost unimaginable. Its main alphabet consists of 65 consonants and 17 vowels, which might look something like:



Tū uonń ruz weahöwelgu\.

(Translates into “*The sound coming from the banks of printers kept on steadily repeating.*”) [7]

Kṣqo'wamentestü liñëkṣ isst^hafûmsîġ.

(Translates into “Aided by the bird’s own stupidity, the man, in inadvertently letting it out of the house, unexpectedly and accidentally killed it without even realizing he’d done so.”) [7]

Under the Sapir-Whorf hypothesis, it is debated whether language forms our perception of reality and language influences our thought process [8]. If this were to be true, languages like *Ithkuil* could hypothetically make its speaker think exponentially faster than one using natural languages such as English, as many concepts could be synthesized into much smaller fragments of thought, compounded in dense formulations. The theoretical gain in efficiency is undeniable under these circumstances, but of course this argument is flawed, if only in considering that our first assumption about the hypothesis is true. Moreover, I have not found any evidence showing an augmented cognitive potential of logical language users - perhaps because not many people speak them. It should be noted that Quijada does not speak *Ithkuil* fluently himself, and nobody else is known to speak it either. Nonetheless, I believe these languages are first and foremost interesting for the theoretical framework they provide to reflect on such issues as cognitive potential and language determinism, rather than efficiency and practicality of use.

Property studies : Redundancy and Compression

The next part of this essay will focus on two properties that are closely linked, and characterize both natural and constructed languages in some way or another : redundancy and compression. This section will be a simple non-technical overview, a statement of their presence or absence, accompanied by explanations and their use/non-use in constructed languages.

Redundancy

The Oxford Dictionaries stipulate that something “redundant” (engineering) is “not strictly necessary to functioning but included in case of failure of another component” [10]. At first glance, it seems like this term might apply for trivial, optional components which are part of a greater whole. Conversely, redundancy can be crucial in terms of persistence, resilience and (re)assurance. This could be understood

in almost any field or situation, going from having many locks on one's front door to the various redundancy systems found in server-based computer systems in case of failure or error (N+1, 1+1, etc.). The reason to have a certain degree of persistence in any given system could be to achieve a higher degree of security or stability (as noted in the two previous examples), but also to insist on a feature or another.

The Windows 7 operating system allows one to store personal files in millions of different locations. The upper limit of files and folders that can be created is virtually unlimited (it is not in truth, but seldom reached by most users), yet most Windows 7 users store the vast majority of their personal files in the same very few folders (Desktop, My Documents, Pictures, etc.). A plausible explanation could be that Windows, in its design, uses a redundant scheme of repetitive promotion and heightened visibility of these folders, to encourage the user to save his or her files in these locations. Although the operating system could do this only once after the initial installation, this feature is persistent throughout its use. Automatic 'shortcuts' to these older and created automatically, default destinations for downloaded files from the Internet are included in those folders, the main work area is "the desktop", etc. The goal of such a feature is surely to encourage a user to keep his or her files organized in a centralized cluster of locations, but nonetheless the choice of 'My Documents' as a more pertinent place to store files than 'C:' is an arbitrary, but important one. It is reinforced by an arsenal of design and interactive features devised by Microsoft and its Windows 7 team of engineers.

I should mention that Apple's operating systems and Linux-based ones have similar strategies for organizing user's files, but this is not the main point. Rather, it is interesting to observe similar features of 'persistence' and 'insistence' in natural languages to emphasize certain ideas, concepts, etc. If one considers the sentence "We had a great night yesterday, it was a lot of fun.", the fact that "it was a lot of fun" is at least partially included in the idea that the night was great (one assumes that fun was to be had, if the night was great) shows a certain level of redundancy. Once again, the aim in such a case would probably be to give extra weight to the fact that the narrator(s) had a great time on this particular occasion, although it is not strictly necessary to the general understanding of the sentence.

Some would argue that redundancy in language is completely superfluous, as per the following examples [24]:

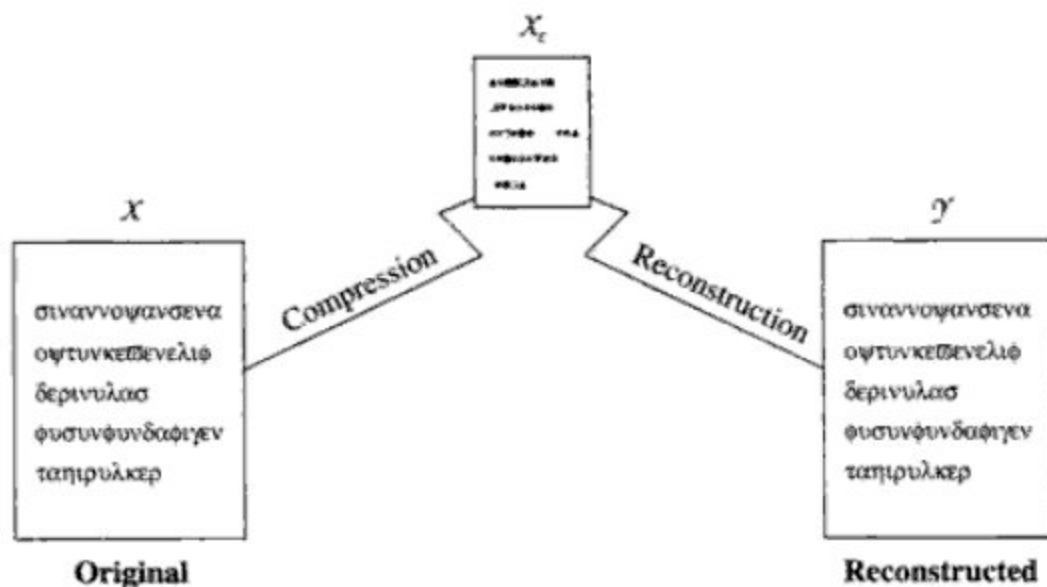
- 1) Needless to say, Mother was not pleased.
- 2) Both John and Mary and leaving
- 3) We shall combine three departments into one.

In (1), the words *needless to say* could probably be omitted since "the speaker or writer who uses this phrase usually proceeds to state that which is *needless to say*", in (2) the word *both* is unnecessary as Mary AND John are leaving (this is generally accepted as an idiom thought) and in (3) the word

combine “means to make one of two or more objects” (Berry, 1971, p.110). Berry, in his book *The Most Common Mistakes in English Usage* stipulates that the “presence [of these words] makes the sentence ridiculous”, attributing this evil to the fact that “1) they are in common use in both spoken and written English 2) they have become acceptable as idioms.” (1971, p.111)

I cannot quite agree with the previous statement, as these sentence formulations could be used for stylistic reasons. Effectively, it would also appear normal to read these sentences just as they are rather than their truncated (and grammatically correct) versions, if they appear more familiar to the reader. Furthermore, this rather purist approach to language use might hinder the natural “organic” evolution of it. It’s usually through informal speech and writing that linguistic precedents are set, which are later adopted officially. By discouraging the use of redundant terms, one makes it that much less of a “natural” language, in my view.

Compression



(Sayood, 2000, p.4)

When dealing with compression, it is generally understood that a mechanism of being “squeezed or pressed together or into a smaller space ; [...] to express in a shorter form ; abridge:” [9] is at work. One could describe the motion as follows (complementing the graph) :

There is the compression algorithm that takes an input X and generates a representation of X_c that requires fewer bits and there is a reconstruction algorithm that operates on the compressed representation of X_c to generate the reconstruction Y . (Sayood, 2000, p.4)

While slightly abstract, this definition makes clear that the reconstruction Y is not the same as original X to be compressed, but rather an identical reconstructed copy. The two main reasons to use compression are 1) to save space in regards to storage - physical or digital. This goes just as much for trash in a dump being crammed together, as movie files being compressed to save hard disk space 2) to reduce bandwidth traffic (usually in digital contexts), i.e. for live transmission of audio signals.

Additionally, there are two types of compression : *lossy* and *lossless*. Lossless compression, as its name implies, offers an exact reconstruction of what was compressed in the first place. Therefore, there is no loss of “quality”, data, information, pieces, etc. The fidelity to the original is quite important in certain contexts, notably when dealing with text : consider the difference in meaning when simple words, letters or punctuation are inverted, omitted or added (i.e. “get me a cake” vs “get me a lake”). Other situations may also call for lossless compression, for example in the healthcare sector : using lossy compression to reconstruct x-ray scans might not be ideal for a doctor diagnosing patients on the basis of these images.

On the other hand, lossy compression occurs when the reconstruction is not identical to the original compressed object. In most cases, this loss of data is predictable/configurable and does not have a great impact on the global understanding of the reconstruction. For example, it is usually considered acceptable to compress images, videos and audio files as the gain in size makes for an interesting compromise, since the lost quality is often not detected (or not enough to be bothering) by our human senses. This concept of fidelity somewhat breaks down when considered in natural languages, as many semantic ‘equivalents’ are difficult to pin down in a perfect manner. Can “a little bit” be replaced by “slightly” in any and all cases? This is a form of compression, but are we losing any of the fidelity in “in a little bit”?

Compression is popular in constructed languages, even more so in engineered languages. Specifically, logical languages often hinge on compression techniques, hoping to condense as much information as possible without losing the essence of a message. As an indicator, *Ithkuil* strives “to allow speakers to say a lot in as few syllables as possible” [15]. Furthermore, semantic entities are broken down into simple phonemes and glyphs, which are compounded to form “sentences”.

As these compression “algorithms” attempt to avoid redundancy, the characteristics proper to the text (if it is to be translated), or the complexity of the message to be transmitted will always dictate to what extent it can be compressed. If, for example, a given text has a rich lexical field, if the message is not particularly clear or has many different implications to consider, these compressed languages might not perform so well. The reason is that many compression techniques depend on grouping similar “objects” together, condensing vague ideas to their essential meaning and replacing entities with smaller/shorter ones. These strategies can work efficiently for large bodies of homogeneous text, but not so much for rich and complex ones.

Conclusion

Within the last few hundred years, no constructed language (including auxiliary ones) has gained enough traction to become remotely threatening to English or Chinese as a universal second language. As it stands, it does not appear that this will happen very soon either. How so? It's a difficult question to answer. Perhaps there is a grounding element, maybe culture, that contextualizes and gives meaning to natural languages in a way that so-called "apolitical" and "culturally neutral" ones cannot fulfill. By doing so, they eliminate the relationship to a place (home?), a time and a history which are tied to the use of language. Perhaps they are also just too difficult to learn, perhaps they are simply not as pleasurable to use as natural languages.

Whatever it may be, I believe the beauty of constructed languages lies in their capacity to open up spaces for us that natural languages cannot. The act of conceiving a whole new toolkit to exchange our thoughts with others is in itself an interesting enough manifestation that makes its exploration worthwhile. Constructed languages allow us to experiment with constraints which are absent or not applicable to natural languages and to operate new modes of communication made difficult by the organic quality of them. Hopefully, this short introduction to constructed languages has made its point clear that their existence serves a vital role in positioning languages in relation to the features we devise and create for them.

Reference list - books

Weinberg, J (2001) An Examination of Logical Positivism, Volume 8, Routledge, London

Brown, K and Ogilvie, S (2008) Concise encyclopedia of languages of the world, Elsevier

Verloren Van Themaat, W. A. (1962) Formalized and Artificial Languages, Synthese, Vol. 14, No. 4, Springer

Berry, T (1971) The most Common Mistakes in English Usage, McGraw-Hill

Sayood, K (2000) Introduction to data compression, Morgan Kaufmann

Reference list - links

[1] http://www.uea.org/info/angle/an_ghisdatigo.html

[2] <http://unesdoc.unesco.org/images/0012/001240/124020e.pdf>

[3] <http://www.carolandray.plus.com/Glosso/Glossopoeia.html>

[4] <http://www.sfwa.org/members/elgin/NativeTongue/ladaanlang.html>

[5] <http://en.tokipona.org/wiki/Category:Philosophy>

[6] <http://www.lojban.org/tiki/la+lojban.+mo>

[7] <http://www.ithkuil.net/ithkuil-intro.htm>

[8] <http://www.nytimes.com/2010/08/29/magazine/29language-t.html>

[9] <http://oxforddictionaries.com/definition/compress>

[10] <http://oxforddictionaries.com/definition/redundant>