1 What Are Constructed Languages?

This book is intended for anyone interested in inventing languages. You don't need a previous background in linguistics or in language invention. Right now, you might not be confident about your ability to invent a language from scratch; but this book will help you reach this goal while learning a lot about languages along the way.

This chapter focuses on what language invention entails. Section 1.1 introduces constructed languages (such as Esperantor or Na'vi) in contrast to natural languages (such as Arabic or Cherokee). Section 1.2 distinguishes constructed languages from creative language forms such as slang and language games. This chapter also covers the main types of constructed languages and the key motivations underlying language invention (Sections 1.3, 1.4). Section 1.5 addresses some important considerations to keep in mind when creating a language, and Sections 1.6 and 1.7 walk you through a fictional scenario and a guided conlanging exercise. Section 1.8 previews the rest of the book, and Section 1.9 suggests additional sources if you want to learn more about the topics covered in this chapter.

Key Words

A priori conlangs
A posteriori conlangs
Alien languages (exolangs/xenolangs)
Alternative languages (altlangs)
Artistic languages (artlangs)
Auxiliary languages (auxlangs)
Compounds
Constructed languages (conlangs)
Engineered languages (engelangs)
Fictional languages
Grammar
  Descriptive grammar
  Prescriptive grammar
Language games (ludlings)
Linguistic systems

Logical languages (loglangs)
Morpheme
Morphology
Naming languages
Natural languages (natlangs)
Philosophical languages
Phonology
Secret languages
Slang
Syntax

1.1 What Are Constructed Languages?

If you were asked to provide an estimate of how many languages are spoken in the world today, what would be your guess? Although it's hard to pinpoint exactly, many researchers calculate that there are between 5,000 and 7,000 languages nowadays. Some, like English or Hindi, are used by millions of people. Others, such as Ainu or Yaaku, only have a handful of speakers. Regardless of the number of speakers, the geographical area where they are spoken and whether they have a writing system, English, Hindi, Ainu and Yaaku are **natural languages (natlangs)** that arose without conscious design. Natlangs spoken today vary greatly; but the rich variation in language they evidence is expanded even more when we consider invented languages, most commonly referred to as **constructed languages** or **conlangs**.

Unlike natlangs, constructed languages are designed consciously. Some are relatively well known, such as Esperanto or Klingon. Some are more obscure, such as Balaibalan or Volapük. Some constructed languages have been around for a while, like Esperanto, invented in the nineteenth century. Others, including Klingon or Dothraki, are more recent. Conlangs appear to be more pervasive now than a few decades ago, but it is important to note that language invention is not an exclusively recent phenomenon. In fact, more than 1,000 constructed languages have been documented since the twelfth century, and the number of conlangs keeps on growing every year.

Do constructed languages resemble natlangs? They do in many respects. All conlangs feature large vocabularies and have developed grammatical systems. The best of them, such as Tolkien's Quenya (High-Elven) and Sindarin (Grey-Elven), also include linguistic irregularities, ubiquitous in natlangs (unfortunately for adult learners). They also take into consideration how languages evolve through time.

Lingua Ignota ('unknown language' in Latin) is often considered the first documented conlang. It was devised by St. Hildegard von Bingen, a Benedictine abbess and polymath who lived in Germany in the twelfth century. St. Hildegard attributed the invention of Lingua Ignota to divine revelation. Lingua Ignota has 1,011 words, rendered in an invented alphabet composed of twenty-three *litterae ignotae* ('unknown letters'). Most of these words are nouns, although there are some adjectives as well; examples are given below. Note that here and throughout, italics are used for words from other languages, and quotation marks for their English translation.

(1) Lingua Ignota words (from Higley 2007:205–230)

a.	*Aigonz*	'God'	d.	*naurizin*	'ring'
b.	*diziama*	'licorice'	e.	*ornalz*	'hair'
c.	*gulzianz*	'potter'	f.	*razil*	'poison'

Having over 1,000 words is impressive, but this alone does not qualify Lingua Ignota as a language. This is because languages are not just collections of words; they are **linguistic systems** involving **grammar**, that is, systematic patterns that define how sounds, parts of words and words can combine in a given language.

Grammar is an often-maligned word, since it tends to be associated to restrictive, often out-of-touch language rules. If you went to school in an English-speaking country, you were probably told at some point that you should not end sentences with a preposition. In this view, common expressions such as *Where (are) you at?* or *What's up?* would not be grammatical. So would a sentence such as *My friend couldn't adopt the kitty she fell in love with*. This use of 'grammar' stands for **prescriptive grammar**. Prescriptive grammar is unavoidable in beginning language courses, or when you are learning to write formally in school. This book, however, focuses on **descriptive grammar**, which describes the ways in which sounds and words are organized in a language, irrespective of whether the language policy (typically self-appointed) decides if it's proper.

Coming back to Lingua Ignota, researcher Sarah Higley shows that in addition to having a sizable vocabulary, this language includes **morphemes** as well. Consider, for example, the similarity in shape and meaning of the Lingua Ignota words in (2).

(2) Lingua Ignota (II) (Higley 2007:102)

a.	*zaimzabuz*	'quince'
b.	*kisanzibuz*	'cotton tree'
c.	*scuanibuz*	'myrtle'
d.	*mizamabuz*	'mulberry tree'

You probably noticed that all words in (2) end in *-buz* and refer to bushes or trees. Because the same meaning patterns with a similar form, *-buz* 'bush, tree' is a **morpheme**, that is, a meaningful unit that cannot be divided further. Thus, since Lingua Ignota displays grammar, it can be considered a constructed language, albeit a primitive one.

Natlangs also have morphemes; consider, for example, the words in (3). The second verb in each example begins with *re-*, which adds the meaning 'again' to the basic verb form; thus *redo* is 'to do again,' *reconsider* is to consider again and so on. We conclude that *re-* is a morpheme in English. You can probably think of many other verbs in English carrying this morpheme.

(3) English verbs

a.	do	redo
b.	consider	reconsider
c.	think	rethink
d.	write	rewrite

DID YOU KNOW?

The Lingua Ignota scholar Sarah Higley is also a language inventor and sci-fi writer under the pen name of Sally Caves. One of the languages she invented is Teonaht, spoken by the Teonim, who live in a region that floats over or submerges below the Caspian and Black seas. Teonaht was recognized with the 'Smiley Award' in 2007. This award, given by conlanger David Peterson between 2006 and 2020, recognized noteworthy conlangs described online during a given year. Other conlangs that received this recognition include Brithenig, Rickchik, Kēlen and Ithkuil. We will consider these and other noteworthy conlangs further throughout this book.

Lingua Ignota features some **morphology** but lacks a well-developed grammar. On the other hand, recent conlangs such as Esperanto, Quenya or Na'vi have fully developed grammatical systems including an inventory of sounds that can be combined in specific ways (**phonology**), a set of morphemes indicating distinctions such as singular vs. plural (**morphology**), and a way in which words combine within a sentence (**syntax**). We will explore the grammatical characteristics of these and other languages throughout this book – some might even serve as inspiration as you work on your own conlang.

1.2 Slang, Secret Languages and Language Games

Conlangs evidence creativity and love of language, but they are not the only outlet humans have to play with or modify language. Throughout history, people have invented and enjoyed **secret languages** and **language games (ludlings)**. Secret languages, codes or 'argots' are used by communities throughout the world to obscure meaning to outsiders and to foment a feeling of inclusiveness for those who are 'in.' Examples of secret languages include Lunfardo, used by criminals in Argentina from the late nineteenth century, and Polari, a secret gay language used in Britain in the 1960s–1970s. Some Polari words are given in (4):

(4) Polari words (Baker 2019:288–296)

a.	*polari*	'the gay language; to talk'
b.	*bona*	'good'
c.	*auntie*	'an older gay man'
d.	*dinarly*	'money'

Slang comprises words and phrases used informally by specific groups of people. It is also a creative form of language which sometimes derives from secret languages, although it emphasizes more the 'insider/outsider' perspective. Slang changes quickly and tends to be used mostly by younger speakers in casual contexts. You probably use slang with your peers; you might have noticed that other age groups use different slang words from you. Some well-known examples of slang include Valley Speak, which originated in California in the eighties, and which still lives on in expressions such as 'whatever' or 'totally.' Examples of fictional slang include Nadsat in *A Clockwork Orange*, and Slayer Slang in the *Buffy the Vampire Slayer* show.

Slang can overlap with language games, as in French 'Verlan,' based on syllable reversal ('verlan' is the reversed form of the French word *l'envers* 'reverse'; the 's' is silent). As shown in (5), in Verlan monosyllabic words, the consonant and vowel are switched. In longer words, syllables are reversed.

(5) Verlan words

	Verlan	French	English translation
a.	*ouf*	fou	'mad'
b.	*looc*	cool	'cool'
c.	*céfran*	français	'French'
d.	*féca*	café	'coffee'

Language games also have secrecy and in-grouping as goals, but they tend to be used more for fun. They occur throughout the world and are particularly favored by children. If you grew up in an English-speaking country, you're probably familiar with Pig Latin. If you spoke Spanish as a child, you might have used or heard Jerizonga. Examples for both are given in (6). Can you figure out how they work?

(6) Pig Latin and Jerizonga words

	English word	Pig Latin	Spanish word	Jerizonga
a.	'sun'	unsay	*sol*	solpo
b.	'world'	orldway	*mundo*	munpudopo
c.	'night sky'	ightnay yskay	*firmamento*	firpimapamenpetopo
d.	'pulverize'	ulverizepay	*pulverizar*	pulpuveperipizarpa

Pig Latin and Jerizonga use specific 'rules' or patterns to obscure words. In Pig Latin, the word-initial consonant is moved to the end, followed by 'ay.' In Jerizonga, each syllable is followed by 'p' and a vowel identical to the one before.

Language games occur in other languages. For example, in Japanese Babigo the syllables 'ba, bi, bo, bu, be' are inserted after each syllable in the word (thus, the word *sushi* is pronounced 'subushibi'), and in Swedish Rövarspråket ('robber language'), each consonant is doubled and an 'o' is inserted between them (transforming *Ikea* into *Ikokea*, for example).

Are language games, secret languages and slang similar to conlangs? Like them, they are creative and evidence conscious design. Crucially, however, they lack a grammatical system different from the language they are associated with. Rather, they modify the ways that sounds and/or words are combined in a natlang to be playful or/and to make the meaning obscure to outsiders. Slang, secret languages and language games are certainly part of 'language play,' like puns and invented scripts, but they are not full languages.

1.3 Types of Conlangs

Several conlang types can be distinguished, beginning with **naming languages**, that is, conlangs consisting mostly of a list of words with little or no grammar. Examples of naming languages include Lapine, the rabbit language in Richard Adam's novel *Watership Down*, and the **alien languages** Runa and Jana'ata in Mary Doria Russell's novels *The Sparrow* and *Children of God*. Most scholars also consider Lingua Ignota a naming language as well since it has little morphology and no syntax.

Auxiliary languages (auxlangs) are conlangs designed to serve as common languages for people from diverse language backgrounds; they combine vocabulary and grammar from two or more languages. Some examples include Esperanto, based on Romance, Germanic and Slavic languages; Afrihili, building on Swahili, Akan and other African languages; and Guosa, combining aspects from Hausa, Yoruba and Igbo. The most successful auxlang to date is Esperanto. One of the reasons why it continues to be popular is that it is relatively easy to learn, particularly for speakers of Romance languages. Although Esperanto was more popular in the past, it still has a thriving community, particularly in Europe, and even hundreds of native speakers.

Artistic languages (artlangs) are designed for creative purposes. They are often connected to specific fictional groups. This is the case of Klingon, Na'vi and Quenya, developed for some of the fictional groups in *Star Trek*, *Avatar* and *Lord of the Rings*, respectively. Artlangs can also stand alone, as in Trent Pehrson's Idrani. **Fictional languages** are those existing in a fictional world; Dothraki and Láadan are examples, and so is Loxian, designed by Roma Ryan for Enya's albums *Amarantine* and *Dark Sky Island*. **Exolangs** (also referred to as **xenolangs** or **alien conlangs**) are languages used by fictional aliens. Examples include Klingon and Fith, spoken by centauroid sapient marsupials on the planet Fithia.

Engineered languages (engelangs) explore one or more properties of language and thus test the limits of how language works. Examples include Láadan and Kēlen. Láadan, created by Suzette Haden Elgin, explores how a language based on female experience would work. Kēlen, by Sylvia Sotomayor, explores what a language without verbs would be like. **Logical languages (loglangs)** aim to be logical and remove ambiguity from language. Well-known loglangs include Loglan, invented by James Cooke Brown, and its successor Lojban.

Other conlang types include **philosophical languages** and **alternative languages (altlangs)**. Philosophical languages are conlangs that aim at building perfect languages that reflect thought precisely. Philosophical languages had their heyday in the seventeenth century; one example is John Wilkinson's Philosophical Language. Altlangs are conlangs set in an alternative history. One example is Brithenig, invented by Andrew Smith; it explores what a Romance language in the British Isles would be like if it had displaced Celtic and undergone Celtic historical changes.

A conlang can belong to more than one category above. For example, John Quijada's Ithkuil is both an engelang and a philosophical language; Láadan is both an engelang and a fictional language (since it features in Elgin's *Native Trilogy* novels), and Heptapod B in Denis Villeneuve's movie *Arrival* can be considered an artlang, exolang and engelang.

It is also useful to distinguish between **a priori** and **a posteriori conlangs**. A priori conlangs are created from scratch, with no direct connection to other languages; examples include Láadan and Na'vi. A posteriori conlangs, on the other hand, are based on one or more languages. For example, Brithenig is based on Latin and Celtic languages, and Eskayan is based on Boholano, Spanish and English.

1.4 Why Do People Invent Languages?

Conlang classification takes into consideration the goals of language invention (Figure 1.1). Some conlangs aim to improve language and achieve a perfect connection with thought (philosophical languages); others have as goals to achieve international (or intercultural) communication (auxlangs). Some conlangs explore the limits of language or linguistic avenues not attested in natlangs (engelangs), while others are designed for artistic purposes (artlang) and/or to enrich a fictional world (fictional languages).

The paragraph above summarizes some of the main reasons why people throughout history, all around the world, decide or have decided to spend weeks, months or their whole lives to developing conlangs. But there are also other possible motivations underlying language invention, including pure fun or enjoyment.

In his 1931 talk 'The Secret Vice,' J. R. R. Tolkien comes out of the language inventor closet and addresses why he devoted so much time and effort to the invention of the languages featured in his literary works. Tolkien highlights artistic pleasure as his primary motivation. In fact, Tolkien wrote *The*

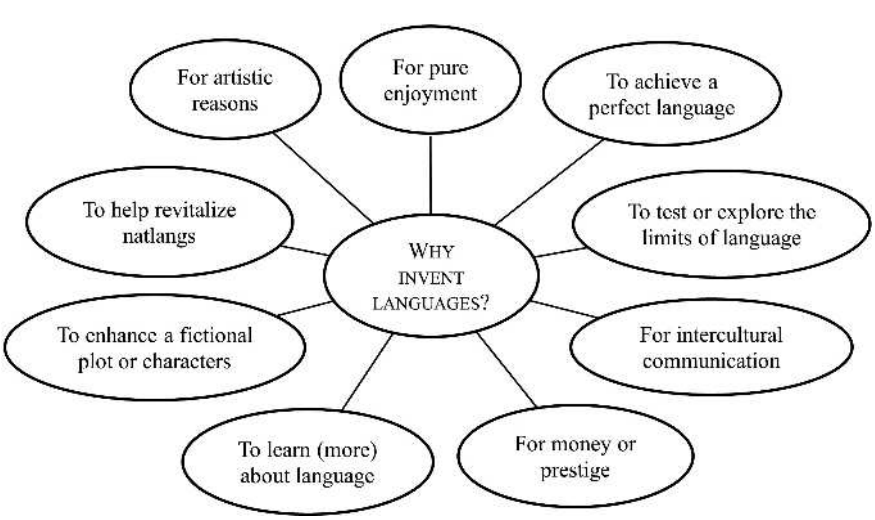

Figure 1.1 Conlanging motivations.

Hobbit, *Lord of the Rings* and *The Silmarillion* to provide a fictional world for his constructed languages – not the other way around.

DID YOU KNOW?

J. R. R. Tolkien was the first language inventor that incorporated linguistic irregularities and historical change in conlanging. Partly because of this, Tolkien is considered by many to be the first modern language inventor.

Some inventors create languages for other reasons, including fame and fortune. Two examples are Edward Rulloff (1819/1820–1871) and Charles Bliss (1897–1985). Rulloff funded his unrelenting efforts to develop a philosophical language through a notorious criminal career involving theft, fraud and even murder. Charles Bliss, the inventor of Blissymbolics, hoped to become famous in the academic world. When Blissymbolics was successfully adopted as a means of communication by the Ontario Crippled Children's Centre in the 1970s, Bliss sued the school. A settlement was reached, which Bliss used to print his own Blissymbolics teaching manual, all to become more famous.

Language invention can also be used to help revitalize declining or dormant languages. Some scholars consider Cornish, Hawaiian, Modern Hebrew and Maori to be invented in a way, since they involve conscious design of vocabulary and grammar, at least to some extent. More obvious examples include Patxohã and Houma. Patxohã is based on Pataxó, a language formerly spoken in Brazil. The revitalization of Patxohã is ongoing; much of its vocabulary was invented by the community, while its grammar is based on Portuguese and Maxakalí. For the dormant Muskogean language Houma, which appears to have been most closely related to Choctaw, there has been a community effort to reconstruct the language since 2013. Even communities that regularly use a language for everyday purposes regularly augment it by inventing new words in response to a changing world, as we will see in Chapter 16. Thus, the distinction between natlang and conlang is not absolute.

It is important to note that there might be more than one motivation at play in the invention of a given conlang. For example, Láadan was invented to explore what a female-centered language would be like; and to incorporate it into a fictional story. In addition, professional conlangers can be commissioned to invent a language for money; their conlangs certainly enrich the fictional world associated with them.

Some people think that language inventors are geeky; some might discount language invention as a waste of time. But as the documentary

Conlanguing: The Art of Crafting Tongues shows, not only is language inventing an old endeavor in the history of humankind, it is also harmless (unless funded by a career in crime, à la Rulloff) and can be extremely fulfilling.

There is one more possible reason to invent a language: to learn about how languages work. This is the main goal of this book: to serve as a guide to conlanging, while learning more about the languages of the world.

1.5 How to Go About Constructing a Language

This book provides step-by-step guidance in inventing an original conlang. Regardless of your goals and motivation, it's important (i) to avoid replicating languages you are familiar with; (ii) to be aware of which characteristics are unique, and which relatively common, in natlangs; and (iii) to strive to be consistent as you build your language from the bottom up.

It is extremely important to be aware of the linguistic characteristics of your native language so that you do not create a simplified or alternative version of it. This is a common pitfall for first-time language inventors. It also happened to me the first time I tried to invent a language. When I was twelve, inspired by Tolkien's Elvish languages in *Lord of the Rings*, I invented a language that, in hindsight, sounded somewhat like Hawaiian as it was based on Spanish but had a limited number of consonants and vowels. I tried to teach this language to my three best friends at the time under the pretense that the radio was broadcasting lessons in it. Because they were good friends, they put up with my attempts to teach them my conlang for a couple of weeks.

You might be multilingual or familiar with several natlangs (and perhaps some conlangs as well). In this case, you might be tempted to model your conlang on another language. But although natlangs and conlangs certainly can serve as inspiration, it is important not to be overly influenced by them, unless of course the conlang you are developing is a posteriori.

It is also important to learn about how natlangs work; specifically, which linguistic characteristics are relatively common in the thousands of languages currently spoken, and which are unique or relatively uncommon. This is crucial to make informed decisions about what you would like your conlang to be like. In general, while conlangs strive for originality and tend to feature unique linguistic patterns, auxlangs typically incorporate well-represented linguistic characteristics, as a compromise among different **linguistic systems** and vocabularies. The best auxlangs are also easy to learn, at least for the speakers of the languages they are based on.

Finally, it is essential to keep track of your choices as you work though designing a sound system, applying it to make up dozens or hundreds of words in your conlang, and developing the morphological and syntactic