

# The Language Construction Kit



Mark Rosenfelder

---

# The Language Construction Kit

by Mark Rosenfelder

⌘ Ⓙ YONAGU BOOKS  
[www.yonagu.com](http://www.yonagu.com) • Chicago • 2010

---

© 2010 by Mark Rosenfelder.

All rights reserved, including the right to reproduce this book or portions thereof in any form whatsoever, except for review purposes.

Kindle edition 2011



# Contents

---

Introduction	4
A naming language	11
The overall process	16
Sounds	29
Word building	58
Grammar	67
Semantics	109
Pragmatics	143
Language families	182
Writing systems	216
Kebreni	237
Word lists	295
More to read	306

# Introduction

---

This book is intended for anyone who wants to create artificial languages— for a fantasy or an alien world, as a hobby, as an interlanguage. It presents linguistically sound methods for creating naturalistic languages, which can be reversed to create non-naturalistic languages. It suggests further reading for those who want to know more, and shortcuts for those who want to know less.



The above is a sample of a constructed language (CONLANG) of my own, Verdurian. If you're curious, it reads, **Ďitelán mu cum pén vead'en er mēsan so Sannam**, meaning “Go forth in peace to love and serve the Lord.” The letter **Ď/d'** sounds like the *th* in *then*; the vowels are similar to those of Spanish. Before I could write this little inscription I had to:

- Decide on the sounds of the language
- Create the lexicon
- Create the grammar
- Design an alphabet
- Modify the alphabet for cursive handwriting
- Translate the text

We'll cover all of this in the book.

The order of the steps above is significant. Working backwards (e.g. creating a text and then devising a grammar to match) will lead to an inconsistent if not incoherent work. A bad example is Hergé's Syldavian; since he basically made it up in pieces, as he needed it, it's impossible to create consistent phonology or morphology for it, based on the scraps of the language provided in the Tintin books.

You can also use this book as a backhanded introduction to LINGUISTICS. To make an artificial language, after all, you have to know a lot about real languages.

## For the impatient

Some books, especially the ones sold in airports, have a lot of padding... the author had an idea for an essay and wrote it as a book.

Not this one. We have a lot of material to cover, and to any expert the presentation will look pretty damn breezy.

However, I've tried to put the most important information first, overall and within each chapter. For instance, the first chapter tells you all you need to know to make a naming language, suitable for filling out a map or making character names. You don't have to be an expert on all this stuff. Well, unless your professor assigned it, in which case, yes, *everything* will be on the final.

# Using this book

---

## Overall structure

We'll start with some overviews. The next chapter is on how to make a very basic language for creating names. After that, we'll look at the process for creating a more detailed language.

The meat of the book is a series of chapters that expand on the major parts of a grammar: sound, word building, syntax, semantics, pragmatics.

After that are some special topics:

- How to create a set of related languages
- Devising writing systems

I've included an annotated grammar of one of my conlangs, Kebreni. The grammar itself serves as a model, but I've also discussed some of my linguistic sources, why I made the choices I did, and what I'd do differently today.

Finally there are some resources: some basic wordlists, and a bibliography.

## But how do I start?

This seems to be a poser for some people. I suppose it's the conlanging equivalent of a writer staring in desperation at a blank page.

Well, first, read the book. You won't be using all the information at once, but it's good to scout out the territory and know what's there so you can go back to it when you need it.

Create an outline for your grammar (see the naming language chapter for an example). Now you can go through the book again and pick alternatives that appeal to you. *There aren't any wrong choices.*

You want a language with ejectives, tones, SOV order, and five cases? Sure, why not?

Think of it as inventing a cartoon character. You choose a particular type of eye, a size and shape of nose, a body shape. The creativity comes in the way the character fits together—it doesn't consist of creating *mind-boggling new organs*.

## Don't read this sentence!

Oops, too late. Don't read *this* one, then.

Anyway, the point is, you don't have to read the book straight through. When you're reading new material, you can absorb it up to a point—then your brain glazes over and it's not really registering any more.

That's fine—just go back to that section later. You can also try alternating reading with creating: read a chapter, then try applying it to your conlang.

## What's it for?

Think about what the language is for.

- Naming people and places. For this you need a brief outline and a wordlist.
- Translating small or large texts, or simply creating the language as an artistic creation—*ARTLANG*. For this you'll need a full grammar; its length will depend on how complete the language is. Translating a text, by the way, is a great way to find what bits of the grammar are

missing.

- ~~An auxiliary language or AUXLANG. Here your choices should be determined by simplicity and accessibility.~~
- Exploration of a logical concept— e.g. you'd like a language with as few words as possible; you want to organize the lexicon scientifically, or it's going to be all gestures. Obviously such experiments aren't constrained by natural languages, but it's still useful to know what's out there, and you should be aware of assumptions that are built into your native language.

## Typographical conventions

I've put technical terms in SMALL CAPS. This tells you two things:

- I didn't make the term up, so you can safely use it in your grammar.
- You can get more information by Googling. (If you get a choice between (say) Linguist List and Wikipedia, look at the former. Wikipedia isn't bad at basic linguistic terms, but it can be wrong or misleading.)

*Italics* are used when I'm discussing a word or phrase. If it's not in English I'll provide a gloss in 'single quotes'.

Example sentences are provided with both a word-for-word gloss and a free translation. I recommend this for your grammars as well— it makes it much easier to see the structure of the language.

**U menya i u nyevo nyet khlyeba.**

by 1s.GEN and by 3sm.GEN not bread-GEN

*He and I don't have any bread.*

Hyphens are used when a word can be separated into pieces: **khlyeb-** is "bread", **-a** is the singular genitive suffix. Periods are used when the meanings can't be separated: e.g. **menya** is a 1<sup>st</sup> person singular genitive pronoun and can't be divided up.

When we get to phonetics:

**boldface** refers to a letter, or informally to a sound

// slashes indicate a phonemic representation

[ ] brackets are used for phonetic representations

### \* *Sidebars*

*Sidebars give tips, warnings, or interesting facts, or provide pointers to additional reading.*

## Your brilliant conlanging career

Conlangers sometimes follow a predictable trajectory.

- First— perhaps before knowing a thing about other languages— they play with their native language— perhaps something as simple as respelling it or creating a new alphabet. Or they create a badass fantasy language that looks like Elvish or Dwarvish.
- Then, depending on personal tastes, they re-invent Esperanto or create a Romance language.
- After learning some linguistics they create a KITCHEN SINK language— one which contains every linguistic feature they've heard of.

It's a natural process and nothing to worry about. But it stands to reason that your efforts will improve over time. So instead of making the first language you do your main conculture's speech, maybe you should work on its ancestor, or a neighboring language.



I wish I'd done it that way myself! My best known and best developed language is Verdurian, but I think most of my later languages are much better done.

---

## **Web resources**

There are a lot of great resources on the web. But URLs rot quickly, so instead of listing sites in the book, I'll list a single permanent URL that will be kept up-to-date with a list of links:

<http://www.zompist.com/resources/>

Kindle users should note that the IPA chart from the print edition, which would be unreadable on the Kindle, is available on this page.

## **Acknowledgements**

Thanks to everyone who read or recommended the web LCK, or provided corrections. I'm sorry I didn't keep names at the time, but Ivan Derzhanski stands out in memory, and John Lawler was an early booster.

Thanks to Daniel von Brighoff and Jeff Burke for reviewing the book, and to Jeffrey Henning and my wife Lida for pushing to get it done. Thanks to everyone on the ZBB for turning this isolated hobby into a community.

*Mark Rosenfelder*

*March 2010*

# A naming language

---

The size of this book may be intimidating— *I have to know all that?* But conlangs don't all have to be huge. For instance, you might just need some names on a map, or an inscription or two. For that, a naming language will do.

An example is Tolkien's Black Speech, of which we have little more than the famous fragment

**Ash nazg durbatulûk, ash nazg gimbatul, ash nazg thrakatulûk agh burzum-ishi krimpatul**  
*One Ring to rule them all, One Ring to find them, One Ring to bring them all and in the Darkne  
bind them.*

Short as it is, this text allows us to learn quite a bit about the grammar of the Black Speech. We can start by noting that the sounds seem similar to English, except for **gh** and whatever **û** represents. These are characteristics that, for English speakers, convey an alien, nasty flavor:

- Many CLOSED syllables (that is, ones that end in a consonant).
- Plenty of STOPS; compare LIQUID-heavy Elvish words like *Lothlorien*.
- Many voiced stops in particular. Final *-zg* looks very alien, though it's just a voiced *-sk* and *-g* occurs naturally in English (as in *he's gone*).

How about the meaning? We know the word *nazgûl*, so we know that *nazg* means 'ring'. Given that, we can create a gloss (a word by word translation):

*ash nazg* one ring  
***durbatulûk*** rule them all  
*ash nazg* one ring  
***gimbatul*** find them  
*ash nazg* one ring  
***thrakatulûk*** bring them all  
*agh* and  
*burzum-ishi* in the darkness  
***krimpatul*** bind them

Let's look closer at the verbs, which I've highlighted. The two verbs which have "them all" in the glosses share a common element *atulûk*. The two that just have "them" share the element *atul*. So we know at least two bits of verbal morphology:

*-atul-* them  
*-ûk* all

Now, it's quite possible that *-atul-* is actually two elements— say, *-at* + *-ul*. Since the examples are all infinitives, perhaps the Black Speech, like English, explicitly marks infinitives. Let's guess that it does, so we have:

*-at-* infinitive  
*-ul-* them  
*-ûk* all

How about *burzum-ishi*? Later on we have the name *Lugbúrz* for Sauron's Dark Tower, so *burz* means 'dark'. *-um* might be a nominalizer (like *-ness*) or a marker of definiteness. So *-ishi* must be 'in', which we must call a POSTPOSITION rather than a preposition since it follows the noun.

## A mini-grammar

Putting it all together, we have a mini-grammar like this:

## **Sounds**

*Consonants: b t d k g z sh th gh m n r l ...*

*Vowels: i a u û ...*

## **Morphology**

*-at- infinitive*

*-ul- them*

*-ûk all*

*-um abstract nominalizer*

*-ul person nominalizer*

## **Syntax**

*numerals before nouns*

*adjectives after nouns (at least in compounds)*

*postpositions*

*subjects before verbs*

## **Lexicon**

*agh and*

*ash one*

*burz dark*

*durb- rule*

*gimb- find*

*ishi in*

*krimp- bind*

*lug tower*

*nazg ring*

*thrak- bring*

If you're writing a naming language, I recommend creating a mini-grammar like this, so that even if there isn't much material, it's all consistent. For instance, if Tolkien needed an object 'me', he would want to remember that 'them' was formed using a suffix.

If Hergé had kept notes like this, he might have kept from creating six different forms of the definite article in Syldavian!

## **Hints for naming languages**

- Note the sounds you use— at the least, those different from English. Keep this list small so it doesn't get cumbersome or contradictory..
- Don't pile on diacritical marks, apostrophes, or odd spellings just to make the words look different— you'll regret it later.
- Start with geographical terms like *city, hill, mountain, forest, lake, river, coast*.
- Add some adjectives that combine easily with these: *big, small, new, high, blue, long*. This immediately gives you a number of names: *New City, Long Lake, Blue Mountain...*
- Different languages don't all work the same! Vary the order of adjective and nouns; use compounds instead of separate words; use different sound systems. Features or people can be named with sentences ("The forest sings") instead of noun phrases ("strong arm").
- If you plan to make the bad guys speak a language with lots of k's and kh's and consonant clusters, you're not the first. Here's a few words from real languages to suggest ways to make

words sound foreign: *Anauá, Neznanovo, Hyōgo, Torbat-e Heydar īyeh, Oaro, Ferkéssédougo*  
~~*Fianarantsoa, Thavung, Yunkunytjatjara, quliuutailat.*~~

- Don't sprinkle the map with a hundred meaningless names. You'll only annoy yourself, perhaps years later, when you want to turn the naming language into a real language. Look at a map of a nation whose language is entirely unfamiliar— perhaps Turkey or Japan. The names aren't just random; they contain repeated elements— very likely the geographical terms and common adjectives mentioned above.

# The overall process

---

The bulk of this book leads you through all the decisions you'll have to make creating a language, and considers the various ways languages deal with these.

Now, you may be thinking, "Do I *hafta*? It's *loooong* and I'm not sure I'm going to *get* all this stuff." So, let's look at it from a very high level.

## Sounds

Even for a simple language, write down all the sounds it uses. Without this, you'll end up with a huge and contradictory set.

Here's a set of sounds (a PHONETIC INVENTORY) to get you started: the Standard Fantasy Language Inventory.

Consonants						Vowels	
	labial	dental	velar	glottal	front	back	
stops	p	t	k		i	u	
	b	d	g		e	o	
fricatives	f	s sh	kh	h		a	
	v	z zh	gh				
nasals	m	n	ng				
liquids		l, r					
semivowels	w	y					

### \* *Um, p t k w t f?*

*The format and labels of this table are explained in the next chapter. This sort of table is appropriate for all languages; English alphabetical order is not!*

Train yourself to check this table when creating words. Want a word *chuth*? Nope, can't do it: *ch* and *th* aren't in the chart. (*But I really want it!* Fine, add them to the chart.)

Don't confuse *sounds* with *letters*. English *sh*, *ch*, *ng* are all single sounds despite being written with two letters (they are digraphs); while *x* is two sounds written with a single letter.

If you want a language for a novel or a game intended for English speakers, it's best to spell the consonants as in English: *sh* instead of *š* or *R* or *sz* or *ch*.

### \* *The trouble with c*

*The sample inventory lacks c. This is intentional: the average reader will read c with English values, generally s before i/e and k elsewhere. If that's what you want, fine, but why not just use s and k? And if it's not, you're just inviting trouble. Tolkien intended c to be always k, with the result that most readers mispronounce names like Celeborn.*

Often it's effective to think in terms of *removing* features from English. For instance, what about making a language without a *p* or an *l*?

If you read nothing else in the long middle of the book, read the chapter on sounds. It's not hard, it's the basis for everything else in linguistics, and you will avoid a lot of silly-looking mistakes if you know something about phonetics.

## Lexicon

Keep a lexicon, as in the Naming Language chapter. Keep it alphabetical in your language, not in English; this will prevent you from reusing words (e.g. using *lug* for 'bring' when you already used

for ‘tower’).

~~If you write on a computer, you can easily search for the English word; if not, maintain a separate English list.~~

Use a column format, like this. You’ll thank me later. (The middle column allows for easy searching for grammatical categories, gives a place later to add special information like declension type, and saves space in the gloss column—you don’t have to write *bear* (v). Plus, tables are easy to convert to HTML.)

**lug** n tower

**How do you make up all the words?** There are several methods.

- Create them one at a time, as you need them (for maps, sample sentences, or texts). This is the best method, since it makes you think about what you’re doing, gives the language a hand-crafted quality, and helps keep you from inventing too many roots.
- Use a word generating tool. This is fine for when inspiration lags, but you’ll probably end up with a formless mess, where you have simple roots for complex concepts (maybe *mopa* for ‘religion’) with no evident relation to anything else (since you picked *naba* for ‘god’ and perhaps even *gupu* for ‘religious’).
- Borrow words from a natlang. Maybe not French. Pick something like Kikongo or Aymara or Malayalam and if anyone notices they’ll probably be flattered.
- Make a proto-language first, then change the words using the Sound Change Applier. This is advanced stuff and we’ll get to it later (p. 190), but it’s actually the easiest, quickest way to create a realistic vocabulary.

You should get in the habit of *doubling up definitions*. That is, instead of making a word for ‘travel,’ make one for ‘travel, voyage, trip’. Similarly, you probably don’t need separate words for

home, house

rod, staff

hole, pit

stream, brook

spurt, gush

happy, content

vast, huge, immense

angry, mad, wrathful

country, nation, land

way, method

soldier, warrior

This will save you work later on when you need one of the alternative words. It also makes your language seem less like a clone of English—its words have their own range of meanings rather than just echoing ours.

On the other hand, be aware of English words that have several very different meanings—you probably shouldn’t duplicate these in your language. For example:

right (direction / correct)

fat (grease / chubby)

earth (soil / planet)

patient (sick person / untiring)

people (persons / ethnic group)

miss (long for / fail to hit)

glass (material / container)

fly (insect / move in air)

fall (autumn / drop)

bear (animal / carry)

**What words do you need?** It depends on what you want the language for. For names and maps, just invent the roots you need. For a reasonably complete language sketch, think about verbs of motion, body parts, kinship terms, simple adjectives, and everyday objects.

At the end of the book there are vocabulary lists with different purposes (p. 295).

## Inflections

English has just a few inflections— plural *-s*, past tense *-ed*, participle *-ing*, and so on. You may be tempted to reproduce these exactly, but there are many alternatives.

As a place to start, many languages have inflections for all PERSONS (*I/you/he*) in singular and plural, e.g. Quechua:

**rimani** I walk

**rimanki** you (singular) walk

**riman** he/she/it walks

**rimanchik** we walk

**rimankichik** you (plural) walk

**rimanku** they walk

An advantage of this system is that you can usually omit the pronouns.

Many languages use inflections where we would use small grammatical words (particles), or entire expressions. For instance:

- modalities (I can, I should)
- aspects of the action: is it completed or still going on, is it repeated
- evidentiality: do I know this for a fact; is it just probable; is it quite unlikely
- definiteness (like our definite articles)
- direction and location (like our prepositions)

The main section of the book will give you many more ideas.

As in other areas, consider *removing* features English has. Do you really need to mark the plural, even tense?

## Derivations

You can multiply the utility of your basic roots, and make your language more consistent, by creating a system of derivations, which might simply be prefixes or suffixes. Some of the most useful:

- person who does
- place
- collection
- tool

- characteristic adjective
- causative
- diminutive

Here's a set of words run through these derivations:

#### Roots

war      book      tree      cow      star

*person*   soldier   librarian   forester   cowboy   astronomer

*place*      battlefield   library   forest   barn   sky

*collection*   campaign   bookshelf   woods   herd   cluster

*tool*      weapon   pen      axe      prod   telescope

*adjective*   warlike   bookish   wooden   bovine   stellar

*causative*   make war   write      plant   stupefy   deify

*diminutive*   fight      pamphlet   sapling   calf      firefly

Five roots turned into 40 words. Naturally you don't have to use all the possible variations; I filled out the whole table just to show how it could be done.

You should find ways of combining arbitrary roots as well, whether through compounding or phraseology. Consider English *battleground*, *skyscraper*, *blackbird*, *mother country*, *tree house*, *dish of the day*.

## Syntax

Syntax includes most of the stuff that most English speakers hardly even realize is part of the language. At the very least, you want to cover the following:

- What's the sentence order: subject / verb / object, as in English? Or something else?
- What's the order within noun phrases? Where do adjectives, articles, and numbers come in relation to the noun?
- Do you have prepositions, postpositions?
- Do you mark cases?
- How are numbers formed and used?
- How do you form questions?
- How do you make negations and negative questions?
- How do you make relative clauses (*The man who stole my name*)?
- What happens when the subject or object of a verb is itself a sentence? (*It's possible he's a liar* / *He claims he's a chicken*).

## Goodies

For extra credit, work out some of the following.

- **Greetings and common expressions.** Because for some reason, people in books can learn another language perfectly *except* for titles, 'yes', 'no', and 'hello', which they insist on saying in their native tongue forevermore.
- The **calendar**: days of the week, years of the month. These can add some flavor to a narrative.



though don't expect readers to actually remember your names. Don't assume that a week  
seven days and that there are 12 months.

---

- The overall **style** used in your language: blunt, polite, pedantic. Do people curse you by saying *You stink* or *May the dust of your ancestors settle in the privy of a diarrheic*?
- Speech patterns that **carry over into English**. If a language doesn't have articles, or the sound, these features could be represented in your translations: *Shtranger is at door, mashte* (Not for the main characters, please.)

## Alphabet

Alphabets are fun to make, and you can use them to write secret notes to yourself. Just a few hints for now:

- Think in terms of *sounds*, not English letters.
- You don't need separate upper and lower case forms. On the other hand, numbers and simple punctuation are useful.
- Don't rush and use the first squiggle you think of for each sound; you'll end up with a dull repetitive alphabet. Doodle a bunch of ideas and pick only the best ones.

For extra credit, think about how your people write— carving in stone, incising clay, brushing on papyrus, writing with brush or pen, gripping a stylus with their tentacles— and create letterforms that reflect that. Carving curved lines is difficult, for instance; and a stylus is more easily used to punch holes or make an impression than to draw.

## Feedback

When I was a lad, it was foolish to expect anyone to be interested in your conlang, except perhaps the future readers of your epic trilogy. And it's still a hobby principally for the self-motivated.

But there are outlets today if you want feedback. You can publish your grammatical sketch on the web, or on one of the websites devoted to discussing or showcasing conlangs.

See the web resources page for places to show your work:

<http://www.zompist.com/resources/>

### \* **Copyright**

*Some people are terribly worried about people stealing their languages. But frankly, unless your conlang accompanies a blockbuster TV show or movie, your problem is going to be getting people to look at it, not having people so excited that they want to plagiarize it.*

*Can you copyright a language? At this point, no one knows— it's never been contested in court.*

*But it's quite clear that you can copyright your work— your grammars, lexicons, and sample texts. If you're American, you get copyright protection even on unpublished work. If you're really worried, you can register your work (even in manuscript form) with the Library of Congress.*

*This isn't to say people won't repost your stuff— there's a billion people connected to the Internet, and some are jerks. Copyright may not help you (the jerk may not even be in your country); polite requests and escalation to their ISP may work better.*

## Models

### Natural and unnatural languages

---

I personally like naturalistic languages, so my invented languages are full of irregularities, quirks, lexical derivations, and interesting idioms.

It's easier to create a "logical" language, and desirable if you want to create an auxiliary interlanguage, à la Esperanto. The danger here is a) creating a system so pristine, so abstract, that it's also impossible to learn; or b) not noticing when you reproduce some illogicality present in the models you're using. (Esperanto actually contains an embarrassing number of irregularities.)

### **Non-Western (or at least non-English) models**

Looking at some non-Indo-European languages, such as Quechua, Chinese, Turkish, Arabic, and Swahili, can be eye-opening.

Learn other languages, if you can. Or just skim a grammar for nice ideas to steal.

- Bernard Comrie's *The World's Major Languages* contains meaty descriptions of fifty languages.
- Anatole Lyovin's *An Introduction to the Languages of the World* surveys all the world's language families, and gives detailed sketches of some important languages Comrie skips.

If you don't know another language well, you're pretty much doomed to produce ciphers of English. Checking out grammars (or this book) can help you avoid duplicating English grammar, and give you some neat ideas to try out; but the real difficulty is in the lexicon. If all you know is English, you tend to duplicate the structure and idioms of the English vocabulary.

#### **\* How do I write lessons?**

*The easiest format for creating a language is a REFERENCE GRAMMAR, such as the grammar for Kebreni in this book. I'd write that first anyway, so you don't confuse yourself. When you're using the language as opposed to learning it, it's useful to have (say) all the information about pronouns in one place.*

*You can follow the format of any language textbook you like. Or look at the Verdurian course on my site.*

*Lessons consist of a reading, new vocabulary, grammatical notes, and exercises, all advancing in difficulty through the book.*

*List all the topics you want to teach, in digestible chunks (say, half a page each), then sort by difficulty and utility. For instance, here's the topics for the first two lessons of my Verdurian course:*

*Basic sentence order*

*Articles*

*Yes-no questions*

*To be (the six basic forms)*

*Possessive pronouns*

*Subject pronouns*

*Gender*

*Verbs (how to form the 3s from the citation form)*

*Double negatives*

*Stuff you can easily live without, like comparatives and causatives, goes near the end. (Or just leave it in the reference grammar.)*

*Difficulty depends on your language, of course— e.g. English adjectives are easy and can be*

*taught right away; Russian or Latin ones have forbidding declensions and should be left till after nouns are mastered.*

---

*The outline can help determine your readings— e.g. you need a reading that introduces basic questions and past tense, or one that focuses on using numbers.*

*Keep new vocabulary to a minimum in each chapter; this can be the hardest thing about writing lessons. Keep a list of the words you've used so far. As you create the readings, re-use words from the list if possible, and if not, add them to the list (and to the vocabulary section for the lesson).*

*You can save a lot of time by making open-ended exercises— e.g. “Continue the conversation between the wizard and the barbarian.”*

## **Am I done yet?**

At some point, after writing four pages of your grammar or forty, you may wonder when a language is done.

One answer you may not want to hear: Never. You can always find something else to write about. There are thousands of books about English, and there are still things we don't know about it.

Or maybe this sounds better: As soon as it meets your needs. You don't need a hundred-page reference grammar to fill out a map. Writing some dialog for a movie might take a careful phonology, a fairly full morphology, and a hasty outline of the commonest sentence types.

I find that a language is fairly presentable after I've created a bunch of sample sentences for the syntax section and about three readings— the process of creating these naturally points out anything I haven't covered yet. For one of my latest languages, Lé, that amounts to about 150 sentences, with a lexicon of about 920 words. It's complete enough that I know I can write texts in the language without having to change the grammar itself much.

Verdurian has a lexicon of over 6000 words. That's enough that I can write quite a bit without having to add to the lexicon... though it's not that hard to find gaps, either.

# Sounds

---

Non-linguists will often start with the alphabet and add a few apostrophes and diacritical marks. The results are likely to be something that looks too much like English, has many more sounds than necessary, and which even the author doesn't know how to pronounce.

You'll get better results the more you know about *phonetics* (the study of the possible sounds in a language) and *phonology* (how sounds are actually used in language).

If you read just one book on linguistics besides this one, make it J.C. Catford's *A Practical Introduction to Phonetics*. Catford goes through the possible sounds systematically, with practical descriptions of how to produce each one even without having heard them.

Linguists use *phones* to refer to a particular sound used in a language.

## Phonetic notation

Language textbooks usually describe sounds by comparison with English, adding recipes for producing unusual sounds. Linguists instead use the IPA (International Phonetic Alphabet), a set of symbols with precise meanings.

I'll ease into using the IPA, since it isn't that helpful till you know something about phonetics. From now on, IPA symbols will be in brackets, and I'll use customary English representations in boldface. E.g. **sh** [ʃ] refers to the English **sh** sound as in *shirt* and tells you that its IPA symbol is [ʃ].

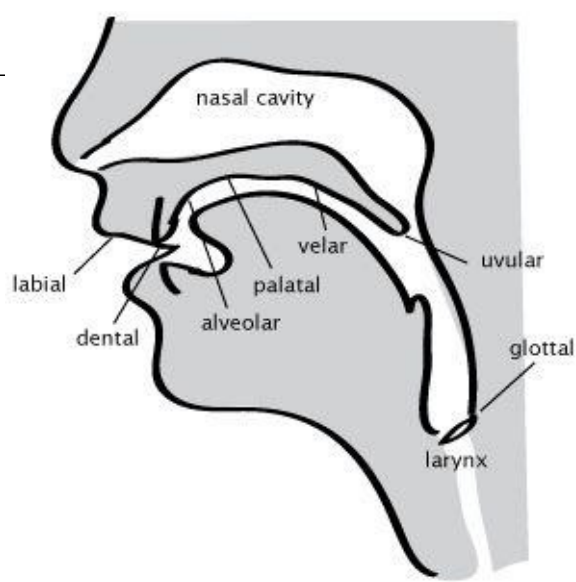
If an IPA symbol isn't given, it's the same as the English representation— e.g. the symbol for **f** is [f]. There's an IPA chart at the back of the book (p. Error: Reference source not found)

## Consonants

You know about vowels and consonants— though the distinction between them isn't as airtight as you heard in school. Consonants can be further organized, however. The most important division is a two-dimensional distinction between *place of articulation* and *degree of closure*.

### Place of articulation

Consonants are formed by obstructing the flow of air from the lungs. The first thing to check is where the obstruction occurs; by convention we start at the lips and move inward.



- labial: lips alone (**w**)
- labiodental: lips and teeth together (**f, v**)
- dental: tongue against teeth (**th**– unvoiced [θ], voiced [ð], French or Spanish **t, d**)
- alveolar: tongue behind the teeth (**s, z**, English **t, d**, Spanish **r**)
- palato-alveolar: tongue further back from the teeth (**sh** [ʃ], American **r** [ɹ])
- palatal: tongue touching the top of the palate (Spanish **ñ** [ɲ], Italian **gn**, Sanskrit **c**)
- velar: back of the tongue against the back of the mouth (**k, g, ng** [ŋ])
- uvular: tongue compressing way back in the mouth (Arabic **q**, French **r**). To pronounce an uvular **q**, pronounce a series of **k**'s while sliding your tongue back as far as it will go. You'll notice a difference in pitch: **q** is about an octave lower than **k**. The **k** in *milk* is part of the way the **q** is pronounced (compare *Mick*).
- glottal: constricting the throat (**h**, glottal stop [ʔ] as in John Lennon saying *bottle*).

The Roman alphabet doesn't have enough symbols, so languages are forced to use letters ambiguously, or use digraphs or diacritics. English and French **t** aren't the same sound, for instance. French **t** is formed by touching the tongue to the teeth, English **t** by touching the alveolar ridge behind the teeth.

## Degree of closure

Consonants also vary depending on *how much* they obstruct the airflow.

- Stops (also called plosives) stop it entirely: **p t k b d g**. In the middle of a word, as in *happy*, this is so fast that we're hardly conscious of the closure. The stop can be lengthened, however, and then we can see there's actually a brief silence while the airflow is stopped. Compare *back* *Beckett*.
- Fricatives just impede the airflow, creating a noticeable hissing sound: **f s sh** [ʃ] **kh** [x]. A fricative can be prolonged indefinitely.
- Affricates consist of a stop releasing into a fricative at the same place of articulation, such as **t + sh** in *tsetse*. English **ch** is actually an affricate, consisting of **t + sh** [tʃ]; likewise **j** is **d + zh** [dʒ].
- approximants impede the airflow only slightly; there's no hissing sound, only a slight change in sound quality: **r l w y**.

Confusingly, the IPA for **y** is [j]. Don't mistake this for English **j**.

- If the airflow isn't obstructed at all, what you have is a vowel.

## More distinctions

Consonants can be voiced or unvoiced; voicing just means letting the vocal cords vibrate.

Unvoiced and voiced consonants usually come in pairs: **p/b**, **t/d**, **k/g**, **f/v**, **sh/zh**, and so on. Sometimes there are gaps:

- German has unvoiced **kh** [x] (which it spells **ch**) but not voiced **gh** [ɣ] (which however exists in Dutch).
- Spanish has unvoiced **s** but not voiced **z**.
- Standard Arabic has voiced **b** but not unvoiced **p**.
- Often approximants only appear in voiced form. Nonetheless it's possible to have an unvoiced **l w y**. For some English speakers, *wh* is pronounced as an unvoiced **w**[ɰ].

Vowels are normally voiced; we'll see some exceptions later.

Voicing isn't entirely binary; languages can differ in voicing onset time (VOT), which is when the voicing starts. English has relatively late VOT— we start voicing initial **b**, **d**, **g**, **j** pretty late; French, by contrast, has early VOT. English also tends to stop voicing pretty early if the consonant ends a word. We really distinguish “voiced” consonants at the beginning and end of the word by other cues.

### Nasalization

Instead of simply stopping the airflow, we can re-route it through the nose, producing nasal consonants **m n ŋ** [ŋ].

The mouth does the exact thing for **b** as for **m**; the difference is that the nasal passage is open for **m**. Thus we call **m** a nasal stop, or just a nasal, with a labial place of articulation.

Similarly, **n** is a nasal dental or dental-alveolar, and **ŋ** [ŋ] is a nasal velar. If a language has other places of articulation, it can have other nasals, e.g. labiodental [ɱ], palatal [ɲ].

### Aspiration

Stops may be released lightly, or with a noticeable puff of air— ASPIRATION. In English, we aspirate unvoiced stops at the beginning of a word (*pot*, *tall*, *cow*), but not after an **s** (*spot*, *stall*, *scow*). French and Spanish doesn't have this initial aspiration (and if you retain it while speaking these languages you'll have a gringo accent).

In Chinese, Hindi, or (Cusco) Quechua, there are separate series of aspirated and unaspirated stops. In Chinese and Quechua, in fact, there isn't a series of voiced stops at all. *Beijing*, for instance, doesn't start with a **b** at all, but an unaspirated **p**.

The IPA symbol is [h]; so the Chinese labial stops are [p p<sup>h</sup>].

### Palatalization

A PALATALIZED consonant is pronounced by raising the tongue toward the top of the mouth. This happens to be about the position for **y**, and a palatalized consonant may sound to an English speaker like **y** if there's a **y** (IPA [j]) after or before it.

Russian and Gaelic have palatalized and unpalatalized versions of most of their consonants. For instance, *да, да* [da da] ‘yes, yes’ sounds very different from *дядя* [dʲadʲə] ‘uncle’.

Palatalization is an example of co-articulation: the consonant is pronounced at (or nearly at) its normal place of articulation, but with the tongue raised. So palatalized [n<sup>j</sup>] isn't quite the same as palatal [ɲ].

## Labialization

A LABIALIZED consonant is pronounced with the lips rounded. For instance, Latin *aqua* ‘water’ was pronounced [ak<sup>w</sup>a] with labialized k. This isn’t the same as the cluster [kw]; with true labialized [k<sup>w</sup>] the lip-rounding is *simultaneous* with the [k].

Any of the stops can be labialized, and fricatives too.

## Glottal games

Most sounds are produced by air moving from the lungs. It’s also possible to create a small amount of airflow by moving the larynx up or down, without any pulmonic airflow at all.

Try it! Touch your Adam’s apple and sing an [a], varying from high to low pitch; you’ll feel the larynx moving. Now do it silently. Finally, keep the vocal cords closed as for [ʔ], put the tongue in [l] position, and raise the larynx suddenly— that should produce an EJECTIVE [k’], a sort of throaty puff.

Now keep your lips closed as for [b], and move the larynx *down* while voicing; this should produce a strangled-sounding [b̥], an IMPLOSIVE. There’s also implosives [ɗ] and [ɟ].

## The consonant grid

Where non-linguists tend to list sounds in alphabetical order, linguists prefer to use a PHONOLOGICAL GRID, with place of articulation across the top, and degree of closure down the sides.

The grid for American English looks like this.

	labial	labio-dental	alveolar	alveolar-palatal	velar	glottal
stops	p b		t d		k g	
fricatives		f v	s z	sh zh		h
affricates				ch j		
approximants	w		r, l	y		
nasals	m		n		ŋ	

Voicing is a third dimension in English; voiced sounds can simply be placed next to the unvoiced equivalent.

This is where the **p t k** order comes from: the stops are listed in order of place of articulation.

## Rhotics

In programming, there’s an aphorism that 10% of the functionality takes 90% of the effort. In phonetics we might say the same about **r** and **l**, which are quite messy.

There are a number of RHOTIC (**r**) sounds:

- American **r** [ɹ] is an approximant, but there are two ways of forming it (which sound about the same):
  - A RETROFLEX **r** [ɻ] is pronounced by curling the tongue up behind the alveolar ridge. Some languages, such as Hindi, have a series of retroflex consonants, such as the stops [ɖ] and [ɗ].
  - A BUNCHED **r** is pronounced by bunching up the tongue thickly under the palate; the tip is drawn back into the body of the tongue.
- In much of England **r** is a post-alveolar approximant— like the retroflex **r** but the tongue pointed at the alveolar ridge, not curled back. (However, younger speakers seem to be adopting a sound more like a w!)

- Another type of **r** is a TAP [ɾ], where the tongue tip is brushed briefly against the alveolar ridge. This can sound like a **d**; thus “veddy” for “very” in attempts to capture certain accents. Spanish single, non-initial **r** (as in *caro*) is a tap; it’s also common in Scottish English.
- **R** can be TRILLED [r̄], which is like a repeated tap caused by vibrating the tongue against the alveolar ridge. Initial and double **r** in Spanish are trilled (as in *rueda, carro*).
- French **r** is a uvular approximant or trill [ʀ].

Don’t confuse any of these realizations with a *dropped r*— that is, one that’s not there! Many English dialects are NON-RHOTIC, meaning that syllable-final **r** is dropped.

## Laterals

LATERALS (**l** sounds) are so called because they’re made with a closure, like a stop, but leaving an opening at the sides of the tongue for airflow.

- English has two distinct **l** sounds:
  - CLEAR **l** is formed with the closure on the alveolar ridge; it occurs at the beginning of a syllable, as in *Luke*.
  - DARK **l** [ɫ] is formed by retracting the tongue (velarization); it occurs at the end of a syllable, as in *cool*. Velarization can be applied to other consonants as well.

Many languages have a clear **l** in all positions; using a dark **l** in (say) Spanish will mark you as a *anglófono*.

- Russian (among other languages) has a *dental l*, with the tongue touching the teeth. If you want a Slavic accent, make your **l**’s dental.
- Then there’s palatal **l** [ʎ], as in Italian *voglio*. Spanish **ll** used to be pronounced this way, and still is in some dialects.
- If the edges of the tongue are closer to the sides of the mouth, so that there’s a noticeable hissing sound, you have a LATERAL FRICATIVE [ɬ]. Welsh **ll** is an *unvoiced* lateral fricative [ɬ] (You may pause to congratulate yourself that you can now work out a triple-barreled term like that. If you can’t, re-read the consonants section!)

## Phones, phonemes, and allophones

I’ve talked about (say) different realizations of English **l** or **p**. However, we need to be more precise about what this means.

Each language has a set of PHONEMES— classes of sounds (phones) that speakers treat as “the same sound”. By convention,

- **phonemes** appear between **slashes**: /l/, /p/...
- **phones** appear between **brackets**: [l], [p]...

This allows us to be brief and precise. For instance, we say that English /p/ is realized as [p<sup>h</sup>] initially and as [p] elsewhere. (The absence of the <sup>h</sup> indicates a lack of aspiration.) The two phones [p<sup>h</sup>] and [p] are called ALLOPHONES of /p/.

You can think of phonemes as how sounds are represented in the speaker’s mental grammar. Speakers are often quite unaware of allophonic variation. We don’t think of the *p* in *pot* and *spot* as different, though phonetically they are.

## Phonemes vs. letters

Often phonemes correspond to letters, but don’t confuse them; letters are an aspect of writing system



- [\*\*read Neon Wasteland: On Love, Motherhood, and Sex Work in a Rust Belt Town\*\*](#)
- [\*\*download Wheat Belly Total Health:Â The Ultimate Grain-Free Health and Weight-Loss Life Plan\*\*](#)
- [\*\*download online With The Material Witness In The Safehouse\*\*](#)
- [\*\*download Back to Basics in Physiology: Fluids in the Renal and Cardiovascular Systems pdf, azw \(kindle\)\*\*](#)
- [\*\*download online House of Robots \(House of Robots, Book 1\)\*\*](#)
  
- <http://test.markblaustein.com/library/Neon-Wasteland--On-Love--Motherhood--and-Sex-Work-in-a-Rust-Belt-Town.pdf>
- <http://academialanguagebar.com/?ebooks/Le-Pacte-Holcroft.pdf>
- <http://econtact.webschaefer.com/?books/Principles-of-Marriage-and-Family-Ethics.pdf>
- <http://www.netc-bd.com/ebooks/My-Organic-Life--How-a-Pioneering-Chef-Helped-Shape-the-Way-We-Eat-Today.pdf>
- <http://cavalldecartro.highlandagency.es/library/Fistandantilus-Reborn--Dragonlance--Lost-Legends--Book-2--Forgotten-Realms--The-Lost-Gods--Book-2-.pdf>