

# Babel



# Babel

Around the world  
in twenty languages

*Gaston Dorren*



PROFILE BOOKS

First published in Great Britain in 2018 by  
Profile Books, 3 Holford Yard, Bevin Way,  
London, WC1X 9HD  
[www.profilebooks.com](http://www.profilebooks.com)

Copyright © Gaston Dorren, 2018

Typeset in Gentium to a design by Henry Iles.

The moral right of the author has been asserted.

All rights reserved. Without limiting the rights under copyright reserved above,  
no part of this publication may be reproduced, stored or introduced into a  
retrieval system, or transmitted, in any form or by any means (electronic,  
mechanical, photocopying, recording or otherwise), without the prior written  
permission of both the copyright owner and the publisher of this book.

ISBN 978-1781256404  
e-ISBN 978-1782832508

Printed and bound in Great Britain by Clays, Bungay, Suffolk  
on Forest Stewardship Council (mixed sources) certified paper.

*Language is such an intimate possession,  
something that one possesses in the same  
measure that one is possessed by it. Language  
is bound up with the foundations of one's being,  
with memories and emotions, with the subtle  
structures of the worlds in which one lives.*

---

**Alok Rai**, *Hindi Nationalism*

丹青

九

# Contents

|                                                |     |
|------------------------------------------------|-----|
| Introduction .....                             | 9   |
| Twenty languages: half the world               |     |
| <b>20</b> Vietnamese   <i>85 million</i> ..... | 15  |
| Linguistic mountaineering                      |     |
| <b>19</b> Korean   <i>85 million</i> .....     | 35  |
| Sound and sensibility                          |     |
| <b>18</b> Tamil   <i>90 million</i> .....      | 53  |
| A matter of life and death                     |     |
| <b>17</b> Turkish   <i>90 million</i> .....    | 67  |
| Irreparably improved                           |     |
| <b>16</b> Javanese   <i>95 million</i> .....   | 83  |
| Talking up, talking down                       |     |
| <b>15</b> Persian   <i>110 million</i> .....   | 95  |
| Empire builders and construction workers       |     |
| <b>14</b> Punjabi   <i>125 million</i> .....   | 113 |
| The tone is the message                        |     |
| <b>13</b> Japanese   <i>130 million</i> .....  | 127 |
| Linguistic gender apartheid                    |     |
| <b>12</b> Swahili   <i>135 million</i> .....   | 139 |
| Africa's nonchalant multilingualism            |     |
| <b>11</b> German   <i>200 million</i> .....    | 157 |
| An eccentric in Central Europe                 |     |

|                                                    |     |
|----------------------------------------------------|-----|
| <b>10</b> French   250 million .....               | 171 |
| Death to <i>la différence</i>                      |     |
| <b>9</b> Malay   275 million .....                 | 187 |
| The one that won                                   |     |
| <b>8</b> Russian   275 million .....               | 203 |
| On being Indo-European                             |     |
| <b>7</b> Portuguese   275 million .....            | 217 |
| Punching above its weight                          |     |
| <b>6</b> Bengali   275 million .....               | 233 |
| World leaders in <i>abugidas</i>                   |     |
| <b>5</b> Arabic   375 million .....                | 253 |
| A Concise Dictionary of Our Arabic                 |     |
| <b>4</b> Hindi-Urdu   550 million .....            | 269 |
| Always something breaking us in two                |     |
| <b>3</b> Spanish   575 million .....               | 285 |
| ¿ <i>Ser</i> or <i>estar</i> ? That's the question |     |
| <b>2</b> Mandarin   1.3 billion .....              | 299 |
| The mythical Chinese script                        |     |
| <b>2b</b> Japanese revisited .....                 | 316 |
| A writing system lacking in system                 |     |
| <b>1</b> English   1.5 billion .....               | 325 |
| A special lingua franca?                           |     |
| Sources and further reading .....                  | 343 |
| Acknowledgements .....                             | 348 |
| Photo credits .....                                | 350 |
| Index .....                                        | 352 |

# Introduction

## *Twenty languages: half the world*

COUNTING THE WORLD'S LANGUAGES is as difficult as counting colours. There are scores of standardised languages, such as English, French, Russian and Thai; making a tally of them is as easy as counting the colours in a Mondrian painting. But most languages have never been standardised. In many areas, there is only a multitude of local varieties, and deciding where one language ends and the next one begins is as hard as distinguishing individual colours in a Turner painting. So there can be no definitive total. That said, 6,000 is a common estimate for the number of languages spoken and signed in today's world – an average of one for every 1.25 million people. Such amazing diversity – what a Babel we live in!

Or do we? Here's another statistic: with proficiency in just four languages – English, Mandarin, Spanish and Hindi-Urdu – you could smoothly navigate most of the world, without any need for an interpreter. Hindi-Urdu and Mandarin are widely understood in the two most populous countries on earth, Spanish will serve you well in much of the Americas and English is the nearest thing we have to a global linguistic currency. So much for Babel, one might think.

The world's biggest languages, which are the subject of this book, are causing the decline of hundreds, even thousands of smaller ones. This is a tragedy, as smaller languages fall out of use across every continent, wiping out valuable knowledge codified in words, stories and names – Alok Rai's 'subtle structures of the worlds in which

one lives'. At the same time, the dominant languages in themselves represent more linguistic, cultural and historical variety than is commonly realised. The contrast makes *Babel* a bittersweet book: the twenty tongues portrayed herein are delicious and dangerous in equal measure.

Between them, these languages are the mother tongues of no less than half the population of the world. Take second-language speakers into account, and the numbers are much larger still. Again, the figures are debatable, but it's safe to claim that at least 75 per cent of people on this planet are able to communicate in one of the Babel Twenty. A less pertinent but more exact figure would be this: over 90 per cent of humankind live in countries where one or more of the twenty are routinely used by central government.

How have these big languages risen to their current station? Individual stories differ, but most have this in common: they are *lingua francas* – languages that bridge the gap between people with different mother tongues.

Two of *Babel's* *lingua francas* – Swahili and Malay – first thrived as trade languages. Later on, several governments embraced them as languages of administration, but even today they're spoken more as second languages – useful gap-bridgers – than as mother tongues. But the primary creator and carrier of *lingua francas* has always been imperialism – Persian, Portuguese and English all outgrew their cradles in this way. Other Asian languages went through similar episodes: Arabic was spread by the caliphate, Mandarin by successive Chinese dynasties, Turkish by the Ottomans and Vietnamese by the kings and armies of the Việt people. Like Portuguese and English, other European languages also piggybacked on colonial empires. Spanish and French were spread by sea, and Russian overland. Nor has history changed its ways – or so the people of South India feel, as they vehemently resist the advancement of Hindi as the all-Indian language.

I have mentioned thirteen languages so far. The remaining seven are German, Japanese, Javanese, Korean and three languages of South Asia: Bengali, Punjabi and Tamil. To categorise these as *lingua francas* would be a stretch. What they have in common

is that they happen to occupy compact but densely populated regions.

If the Babel Twenty have triumphed in different ways, that's only the beginning of their diversity. Unsurprisingly, all languages are different in the words they offer, the grammar they employ and the sounds they travel on. Their writing systems are not just alluringly varied in appearance, but also profoundly dissimilar in their functioning. People have different collective feelings about their languages: we find veneration, pride, protectiveness and sometimes indifference, but also, especially among second-language speakers, resignation and even loathing. Languages are put to different uses: most, but not all, are popular with governments and businesses; some have long and rich literary traditions, others less so; some will be maintained by migrants for several generations, while others are soon given up. All languages display internal diversity, but the patterns differ: usually, there are regional varieties; sometimes, there's one for speech, another for writing; or one for formal conversations, the other for informal chats; or different varieties for speaking to one's social superiors, inferiors or peers, et cetera. In other words, apart from being a unique system of communication, each of the Babel Twenty also has its own language history and its own linguistic culture. They are worlds unto themselves.

In the following twenty chapters (plus one bonus semi-chapter) I will peer into one of these worlds, starting with the smallest of the bunch and working towards the biggest, the world's 'linguistic superpower'. But while each story will focus on a language, it will also focus on an issue, on one particular feature of that particular language. For instance, what does it actually mean for Russian to be 'related' to English? How do non-alphabetic scripts, such as those of India and China, do the same job as our twenty-six letters? If Belgium and Canada have trouble keeping the linguistic peace, how do multilingual countries such as Indonesia manage? How did tiny but colonial Portugal spawn a major world language – and why didn't the Netherlands? Why do Japanese women talk differently from men? And how did this book gain the author two Vietnamese nieces?

## About Babel – some (im)practicalities

Each chapter begins with a short profile of the language under discussion: its various names, its linguistic pedigree, the numbers of speakers, some basics about grammar, sounds and the writing system as well as information about loanwords (which are the main sources of borrowing and which words English has adopted from this language?). The numbers, of course, are questionable, as language statistics are highly erratic; I've consulted many sources, ignored the implausible outliers, averaged the others and rounded the outcome to the nearest catchy figure.

It's hard to represent the unfamiliar sounds of foreign languages without making use of the International Phonetic Alphabet, which can be opaque for non-specialists. I've tackled the problem in two ways. In most cases, I've tried to approximate the pronunciation of the foreign words by referring to either English spelling conventions (e.g. by using /ee/, as in *bee*) or some 'continental' spelling (where the same sound would be represented as /i/ or even, in passages where vowel length matters, /i/). In a few cases, where this wouldn't work, I've included references to my website ([languagewriter.com](http://languagewriter.com)), where under the menu item 'BABEL', there's a page with sound files.

When citing words, phrases or sentences as examples, I've tried to be typographically consistent in the following way:

**BABELI** Small caps: words in foreign languages. But for legibility, whole sentences are quoted in italics.

*Babel* Italics: English words, but also sentences in other languages. Italics for emphasis are used too.

/baybl/ Slashes: phonetic approximations.

'Babel' Inverted commas: English translations of words.

(Babel) When English translations are placed in parentheses, no inverted commas have been added.

Precisely half of the Babel languages are written in a script other than our Roman alphabet. Words in Russian, Mandarin and so on have been transliterated or transcribed. More than one system is in

circulation for each of several languages, so if you feel that Korean, Japanese or Arabic words in *Babel* look somewhat different from what you were taught, this may be the explanation.

I've taken great pains over the many foreign words cited in this book, but it is nonetheless likely to contain spelling errors, especially in the non-European languages. I would be grateful for any corrections that readers might send me (through the contact page on my website – [languagewriter.com](http://languagewriter.com)); they will inform later editions and translations. Meanwhile, let's appreciate such errors for what they are: evidence that we live in Babel still.

Gaston Dorren, 2018

# 20

## Vietnamese

## TIẾNG VIỆT

**85 million speakers**

Around 75 million native Vietnamese speakers live in Vietnam, where it is the only official language; half a million in Cambodia. Some 2 million Vietnamese live in the US and substantial numbers in France, Australia, Canada, Germany, the Czech Republic and the UK. There are 5 to 10 million second-language speakers, mostly ethnic minorities, in Vietnam.

## Vietnamese

**SELF-DESIGNATION** TIẾNG VIỆT, sometimes TIẾNG VIỆT NAM or VIỆT NGŨ

**FAMILY** Vietnamese is far and away the most widely spoken member of the Austroasiatic family, which also includes Khmer, the national language of Cambodia. More than 100 smaller Austroasiatic languages are spoken by minorities throughout Southeast Asia and eastern India.

**SCRIPT** Roman alphabet, with a remarkable number of diacritics (accents). Until the early twentieth century, a character script based on Chinese was in use.

**GRAMMAR** See main text.

**SOUNDS** See main text.

**LOANWORDS** In prehistory, agricultural terminology was adopted from the neighbouring Tai languages. A huge number of words from Chinese languages were borrowed over a period of some two thousand years, up until the mid-twentieth century; according to estimates, 30 to 60 per cent of Vietnamese vocabulary is of Chinese origin. During French colonial rule, scores of French words were adopted and adapted to the Vietnamese sound system and spelling. Today, English words trickle in. For examples, see main text.

**EXPORTS** English mainly uses Vietnamese words to refer to elements of Vietnamese culture. The best known may be *pho* (PHỞ), a popular broth-based dish. During the Vietnam war, some Vietnamese words temporarily trickled into American military jargon, such as *đi đi* (ĐI ĐI) for 'go away', *Quan Canh* (QUÂN CẢNH) for 'Military Police' and *so mot* (SỐ MỘT) for 'the best', literally 'the number one'.

**ACCENT OBSTACLES** Vietnamese has strong regional variety in terms of pronunciation and vocabulary, though not in grammar. For the language student, this is a major obstacle: it's easy to tell the major accents apart, but more difficult to understand them, especially the ones spoken in the south. Indeed, even northern native speakers have trouble dealing with some of the less familiar southern dialects. The recommended standard pronunciation is a mixed bag of some southern and rather more northern characteristics.

## 20: Vietnamese

### *Linguistic mountaineering*

I'M HALFWAY THROUGH my three-week stay in Hanoi when I hear myself say: '*Bác học tiếng Việt một năm và ruỗi - no, wait: ruỗi - rồi nhưng chưa có thể nói không được!*'

The words come out haltingly, but I see Loan, my teacher, nod, so I may indeed be saying what I'm trying to say: that I've been studying Vietnamese for a year and a half now and I still can't speak it. Have I really managed to get that across? If so, I've just produced my longest - and most paradoxical - Vietnamese utterance so far.

Loan (pronounced 'lwahn') is silent for a second or two. Then she shakes her head. 'Let's correct that.'

I ignore the sinking feeling in my stomach and try to make my answer sound light-hearted: 'Why? You feel I'm wrong in saying that I can't speak the language?'

'The sentence has several errors,' and she starts listing them. 'Now say it again.'

*'Bác học tiếng Việt một năm ruỗi rồi nhưng chưa thể nói được.'*

I like studying with Loan. Not only does she speak good English (though that may be a double-edged sword), but she's also lively, fun and full of enthusiasm for language study, both her students' and her own. We've been to a museum, a bookshop and an eatery together, and she's even introduced me to a few of her friends. But while she's a dedicated teacher, there's no denying that she can be a little stern. Rather than praise what progress I may make, she firmly believes in the motivational force of her disappointment, saying things like, 'Why don't you know that word, *bác* (uncle)? We read it only yesterday!'

'Undoubtedly, but I can't possibly memorise every single word we come across, *cháu* (niece), or even half of them! Learning Vietnamese vocab is *khó khăn* - it's difficult!'

‘You should, though. You’re such a good student!’

I sit up.

‘But it’s true that your conversational skills haven’t improved much.’

I slouch.

‘Yet they have somewhat. And your pronunciation certainly has.’

But not right now, it hasn’t, because the rare compliment leaves me speechless. That’s how we roll, my twenty-year-young Vietnamese teacher and I.

But why am I here in the first place, 6,000 miles away from my friends and family, struggling with a language that, except for the odd expat, only the oddest foreigner ever ventures into? What made me do it? The answer is: this book did.



The Temple of the Jade Mountain in central Hanoi. For centuries, most writing was done in Classical Chinese. An adapted version of its script was also used for Vietnamese.

As I mentioned in the introduction, one would need to learn twenty languages to be able to speak to half the world population in their mother tongues. Some people have actually pulled that off. I once shook hands with the American polyglot Alexander Argüelles, who speaks sixteen or seventeen of the Babel languages and many more that aren't in the top twenty. Personally, I've studied seven of the twenty at some point in my life (English, German, Spanish, French, Portuguese, Russian and Turkish), resulting in levels of familiarity ranging from fluency to (more often) mere acquaintance. The idea of having a go at the full set of twenty did briefly tickle me. But I soon realised that I would never even cover the basics of them all within a time frame that my publisher, my bank account or, indeed, my life expectancy would allow. So I decided to try just one – but a challenging one.

Vietnamese seemed a good choice. Unlike most other options, such as Arabic, Hindi and Korean, it wouldn't require learning a whole new alphabet, nor, heaven forbid, thousands of Chinese characters. And since Vietnamese was going to be the protagonist of the opening chapter, it would allow me to dip into many features of foreign languages early on in the book. An additional, private motive was that I hoped I might address Tuyet, our Vietnamese cleaning lady, in her mother tongue, preferably by surprise. So I made my choice, picked a self-study book that appealed to me\* and set out on my journey. This chapter is an account of that journey.

## Early acquaintance

Written Vietnamese makes a striking first impression: no other language that I know of has so many diacritics. As a young boy, I found my father's *Paris Match* exotic because of the numerous é's, à's and î's and endless apostrophes, yet French looks stark and unadorned compared to Vietnamese. With no fewer than nine different diacritics (á, à, â, ã, ă, â, ã, đ and ơ – I love the informal term 'whiskered o' for this last one), it's a language for people with keen eyes. Words and even letters with several diacritics are by

---

\* I'm grateful to Assimil for providing me with their Vietnamese course. In the 'Sources and further reading' section at the end of this book, I list the learning resources I've used.

no means rare: I soon discover that Tuyet's name is actually spelt Tuyết (and that it means 'snow').

These accents are necessary to ensure correct pronunciation, but they add considerably to the burden of memorisation. Fortunately, written Vietnamese also seems to have a plus side: nearly all words are monosyllabic. Words of up to six letters must be easier to memorise than those centipede-shaped things that German, Russian and Turkish are so fond of.

More good news: unlike English, Vietnamese spelling reflects pronunciation pretty accurately. Not that all spelling rules make intuitive sense. Why have three different ways of writing a /z/ sound – and why should the letter z not be one of them? Also, why have three different ways of writing a /k/ sound? (Though of course, in English, we write *can*, *keen* and *queen* rather than *kan*, *keen* and *kween*.) Another intricacy of Vietnamese is the distinction between *t* and *th*, the former a /t/ sound, the latter a /t<sup>h</sup>/ sound – that is to say, a /t/ with a little puff of air. In English we write both as *t*, so we can hardly complain if Vietnamese prefers to be more precise.

On the whole, I would call Vietnamese pronunciation easy if it weren't for the elephant in the room, the big fat singing elephant that bedevils most East Asian languages: *tone*. Each word must be vocalised according to one 'tune' from a total repertoire of six. I knew this beforehand, but I hoped that as a native speaker of a tonal regional language (Limburgish) I would take it in my stride. No such luck: six tones is very different from the paltry two I'm used to. Use the wrong tone, and 'here' becomes 'there' (ĐÂY, ĐẤY) and the meaning of ĐI changes from 'go' to 'prostitute' (ĐỈ), 'scrotum' or 'ill-treat' (both ĐÌ). Fortunately, Vietnamese spelling is helpful: five of the nine diacritics are there to indicate the correct tone; the sixth tone is marked by the absence of these five. The result of this is that Vietnamese writing is equal parts text and score.

There's something insidious about the grammar of Vietnamese. On the one hand, it's easy in all the places where experience has taught me to expect hard work. Verbal conjugation, the ordeal of most European languages? None to speak of: you just add a handful of simple particles that never change, and even those can

often be omitted. Cases, as in Latin and Greek, the Slavic group and German? Entirely non-existent. Irregular plurals, as in the Slavic group,\* Danish and German? Vietnam doesn't even have plurals as we know them. All of which can be summarised as: no endings! Vietnamese words stay as they are. They never change. Hard to believe, but true.

But just when I'm beginning to hope that Vietnamese might turn out to be rather easier than I thought, several gremlins emerge.

**Gremlin 1:** the personal pronouns, which are ... hell, where do we even start? The trouble with Vietnamese pronouns is that there are so many. You do not simply say 'I' and 'you' – there is a multitude of Is and yous to choose from, partly depending on gender, but also on respect and age. The pronoun you choose creates a specific type of relationship. Even if you use the most neutral word for 'I', *tôi*, it's not *really* neutral, because it makes you sound terribly aloof, which is no way to win friends.

Much more common than the neutral or distant pronouns are pronouns that, in literal translation, designate all sorts of family relationships. If you are somewhat older than me, I will address you as ANH (older brother) or CHỊ (older sister), while referring to myself as EM (younger sibling). However, if you're younger, I will call you EM and refer to myself (being male) as ANH. Note that all three words may mean either 'I' or 'you', depending on who is saying them to whom. If the characters in a Vietnamese dialogue use these pronouns, the first thing to do is figure out, from context, who's junior and who's senior, so you can tell which pronoun refers to whom. Entirely different words again come into play if the age difference is wider. The reason why Loan and I in the above dialogue called each other BÁC and CHÁU (literally: 'uncle/aunt' and 'niece/nephew') is that I'm about the same age as her parents. Incidentally, I didn't insert these words into the dialogue for local colour: Vietnamese people commonly use them even in English.

**The second gremlin** is what I will call 'labyrinthine sentences'. Of course, when you're starting on a new language, it's normal to feel

---

\* Note that the Slavic languages are in several ways the diametrical opposites of Vietnamese. For decades, the Soviet Union was Vietnam's main ally, which drove two suffering generations to studying Russian – which is Slavic.