



Introduction

This is a beach book, a long-flight book, a public-transportation book, a before-I-go-to-bed book. It has been written to entertain first and educate second. What is meant by a geography of words here are language-specific and culture-specific configurations of words. Fascinating cases of differences in the way that languages organize their vocabularies are portrayed here. As famous Russian-American linguist Roman Jakobson once noted, languages differ in what they must convey, not in what they may convey. The examples discussed in this book concern situations where various languages express ideas in their word stocks in a way that may be curious to English speakers and those where English may be strange to speakers of other languages. The choice of English is determined not only by the fact that this book is written in English and intended for English speakers but also that for many years now the English language has held a special place in international communication.

An anecdote involving the aforementioned brilliant linguist and philologist Roman Jakobson is illustrative in this regard. The saying has been circulating among linguists for many years that Jakobson fluently spoke five languages – all in Russian. It is only recently that I learned, thanks to Anatoly Liberman, another brilliant scholar, that the statement is incomplete – the full version, which actually celebrates multilingualism, reads, Jakobson fluently spoke five languages, all in Russian, and Jespersen was fluently silent in ten languages (Jespersen was another prominent mid-twentieth-century linguist). More importantly, I learned that this statement is fake news altogether. Jakobson did have a strong Russian accent in English, but, per Liberman, he spoke French, German, and Czech almost without an accent. This story shows the special place of English. The prominence of English dwarfs other languages, so anything impressive relating to other languages was trumped by Jakobson's nonnative Russian-colored English. This is akin to the inaccurately narrow usage of “bilingual”

in the United States, where it actually means ‘speaking English and Spanish’ and nothing else, or in Canada, where it means ‘speaking English and French’, whereas bilingualism can include any given pair of languages.

Since this text is meant to be short and entertaining to read, no references have been inserted in the body of the text; however, in Part VII, there is a section on recommended readings for each chapter and a list of references. Languages that use the Latin script are given in their original spelling. Those that use other scripts are given in the original and also transcribed using the closest English approximates and in the so-called pronunciation respelling. This kind of transcription does not always reflect all the nuances of the sounds in the original language, but rather gives a general idea of how the word is pronounced using English sounds. Transcription tables for non-Latin scripts and the specific characters of Latin-based scripts are provided in Part VII in the sections titled *Pronunciation Respelling* and *Language-Specific Latin Characters*.

Given that this book is intended for a general audience and is not a scholarly monograph, some common-parlance shortcuts are made in it. When stating that a language has or does something, I am using a shorter way of saying that its speakers, or most of them, commonly have those features or engage in such practice. Needless to say, languages do not do or have anything; they are aggregations of their speakers’ linguistic knowledge, habits, and practices.

The cases of cross-linguistic differences presented here are a testimony to the diversity of the human condition, thought, and its linguistic expression. This richness should be cherished – a world in which one language prevailed would be as boring as a natural world with only one kind of tree and animal. This book is thus meant as a small contribution toward preserving the richness of human linguistic diversity.

The architecture of books can assume various forms. Slovenians, speakers of the Slovene (aka Slovenian) language in south Central Europe, have a useful distinction when talking about series (as in series of stories in a book or TV series/shows). The term *nanizanka*, literally, ‘beaded series’, refers to a series where the plot is separate in each story or episode. By contrast, the term *nadaljevanka*, literally, ‘continuing series’, refers to those where the plot stretches over the whole series. This book is a *nanizanka*, so one can hop from one chapter to another, without the need to follow their order.

As already noted, Jakobson once observed that languages differ not in what they may express but rather in what they have to express. In numerous cases discussed in this book, a peculiar feature of one language can also be rendered in another, but it does not have to be. It is often so that one language has a so-called culture-bound word, which is not rendered in one word in many other languages, although they may have something similar. Thus, the Portuguese language has the word *saudade*, which can be rendered in English as a feeling of longing, melancholy, or nostalgia. The Internet is awash with examples of these culture-bound words from various languages, from Bantu *bilita mpash* ‘an amazing dream, opposite of nightmare’ to Finnish *myötähäpeä* ‘vicarious embarrassment’.

To further illustrate these differences, Russian speakers need to differentiate two nuances of blue and use separate adjectives for them: sky-blue *голубой* (*galooboy*) and deep blue *синий* (*sinii*). This is possible to do in English, but it is not obligatory; one can simply say *blue*. Similarly, Mandarin Chinese has to differentiate between one’s older 哥哥 (*gēgē*) and younger brother 弟弟 (*dìdì*). In English, this is not a compulsory distinction, so the word *brother* can simply be used in both cases.

Part I, *How Words Are Studied*, addresses the issues in the scholarly studies of words. It outlines some key concepts in the study of words for those who would like to look at the cases of global lexical diversity presented in Parts II–V with a more focused lens. The first chapter of this part is titled *What Is a Word?* It tackles questions such as how to differentiate a word from a nonword and what features should be ascribed to a word. The second chapter, *The Internal Affairs of Words*, looks at the inner structure of words, the intricate networks of their senses. Finally, the chapter *The External Affairs of Words* looks into the issues of the links between words in the lexicon and how words cross from one language to another.

Starting from Part II, the chapters present fascinating cases of cross-linguistic variety, each in its subject-matter area. Chapter 1 in *How Words Are Carved Out*, the second part of the book, is titled *1 = 2, 5, 6, or 7*. It focuses on the differences in how languages construe their numbers. Chapter 2, *Beer Eyes and Wine-Dark Sea*, looks into peculiarities in the perception of colors. Kinship terms are discussed in Chapter 3, *Second Cousins Twice Removed*. Chapter 4, *I Have Three Sons and a Child*, looks into the construal of gender in various

languages. Chapter 5, *Concepts on the Chopping Block*, explores cases where languages exhibit differences in carving out their concepts. Chapter 6, *Unripe Bananas and Ripe Tomatoes*, uncovers differences in how tastes are perceived, while Chapter 7, *Mums and Clocks Mean Death*, addresses the connotations around certain objects and then common prototypes. Chapter 8, *The Past Is in Front of Us and the Future Is behind Our Back*, is about the conceptualization of time, while Chapter 9, *Far and Wide, Here and There*, looks into the conceptualization of space. Various metaphors are explored in Chapter 10, *Bottles with Throats*. Chapter 11, *Setting the TV on Fire and Extinguishing It*, looks into the encapsulation of previous periods in the society in some terminologies.

Part III, *How Things Are Done with Words*, is opened by Chapter 12, *Traduttore, Traditore!*, which looks into the glitches in translation and interpretation. Chapter 13, *May You Suffer and Remember*, explores maledictions, while Chapter 14, *I Screw Your 300 Gods*, discusses expletives. Unusual idioms and sayings are explored in Chapter 15, *Either He Is Crazy or His Feet Stink*. The next two chapters explore onomatopoeia – those used in talking to animals are covered in Chapter 16, *Shoo and Scat*. Words used in reference to animals are addressed in Chapter 17, *A Dog and Pony Show*. Words used to imitate surrounding sounds are discussed in Chapter 18, *Blah-Blah-Blah, Yada-Yada-Yada*. This is followed by the discussion of obscenities in Chapter 19, *Acts of Darkness*. Chapter 20, *This for That*, discusses the strategies of referring to concepts for which we do not have words. The next two chapters are about words in planned languages and pidgins, *Me Tarzan, You Jane* (Chapter 21), and the words one is expected to acquire to speak a language, *How Many Languages Do You Speak?* (Chapter 22). The so-called false cognates are addressed in Chapter 23, *Harmful and Shitty People*.

Part IV is titled *How Words Are Born*. Chapter 24, *Cars with Tails and Leadfooted Drivers*, gives a general introduction to this part, discussing the ways in which new words and their meanings emerge. Naming new concepts is explored in Chapter 25, *Monkey, Dog, Worm, Snail, i.e. “Crazy A”*. Chapter 26, *Rovers and Ski-Rolls*, addresses the names for new nonmotorized vehicles. Chapter 27, *Extra Crispy Soccer Players*, looks into the peculiarities of athletic terminology. Chapter 28, *Chinglish and Eurenglish*, discusses words

in hybrid varieties of English. Forms of address are discussed in Chapter 29, *Comrade, Sir*. The story goes on with Chapter 30, *Beer and Whiskey Mightily Risky*, which talks about terms for alcoholic beverages. Chapter 31, *SOFs and SOWs*, discusses acronyms and abbreviations.

Part V, *Where Words Live*, looks more closely into the relationship between language and culture. Chapter 32, *Old-Lady Torturers, Horse Killers, and Bad Mornings*, discusses unusual personal and place names. Chapter 33, *A Fleeing Bus*, talks about different encapsulations of the locus of control. Chapter 34, *I Wish That You Enjoy in What You Have Deserved!*, is about congratulations and responses to them. Chapter 35, *Happy Hunting Ground*, is about the words used to talk about death. Chapter 36, *A Language Is a Dialect with an Army and Navy*, is about the degree of lexical differences and their role in determining if something is a variant of the same language or a separate language.

Part VI, *A Word After*, is a brief reflection on some broader consequences of the global language diversity presented in Parts II–V.

Part VII, *Words about Words*, contains the apparatus of the book, that is, a collection of notes, references, tables, and indexes to accompany the main text. *Selected Chapter Readings* for each chapter are provided first for those who seek more detailed information about the topics discussed in the chapters. A list of *References* is provided next, followed by *Pronunciation Respelling and Language-specific Latin Characters*, which should enable those who seek to pronounce the words used as examples throughout the book. Finally, an *Index of Languages* is provided with references to the pages where they are used and the information where they are spoken and how many native speakers they have, followed by a *Person and Subject Index*.

PART I

How Words Are Studied

An unsolicited piece of advice to the reader (I am a Slav, after all, and that is our national pastime)

If you are not interested in the technical aspects of the study of words, please skip this part and go straight to Part II

A What Is a Word?

This book is primarily about global linguistic diversity in the LEXICON (a collection of words in a language) and LEXICAL SEMANTICS (meanings of these words), seen through the eyes of the linguistic field of LEXICOLOGY (the scholarly study of words). It then makes perfect sense to preface the discussion of concrete cases of lexical and semantic diversity with an introduction of key concepts in the field that are of relevance in the main chapters. Those who are just looking for interesting cases of cross-linguistic differences should skip this chapter, and the next two chapters, and proceed straight to Part II of the book. Only those concepts immediately relevant for the discussion in the main chapters are discussed here. Consequently, Sections A, B, and C, are not an ABC of the linguistic study of words, but rather an ABC of navigation (hopefully, cruising) through this book.

As noted, the field of linguistics that studies words is called LEXICOLOGY, not to be confused with LEXICOGRAPHY, which is the art and craft of dictionary making. The vocabulary items that dictionaries define are often one word long but can also be multi-word items. The technical term that covers both is LEXEME or LEXICAL UNIT (not out of sheer whim, but because there are one-word lexemes, such as *break* and *spring*, and also multi-word lexemes, such as *spring break*). The level of the language system that is formed by lexemes is called the LEXICON, a level distinct from that formed by the sounds, called PHONOLOGY or that formed by combination of words, called SYNTAX.

Lexemes and lexicons do not hang in the air; they belong to their respective languages and cultures. One of myriad definitions of culture was proposed by Hofstede and Hofstede, “[Culture] is the collective programming of the mind which distinguishes the members of one group or category of people from another.” (see Part VII for this and all further references). Cultures around the world are very different, and scholars of cultural anthropology have identified the parameters of cross-cultural variation. Two of these researchers are of relevance here, Hall and Hofstede.

Edward T. Hall proposed the so-called classic patterns. The first is the opposition between **MONOCHRONIC** and **POLYCHRONIC** cultures. The former cultures perceive time as one line with past, present, and future separated. In the latter cultures, past, present, and future mix with each other on multiple timelines. As a consequence, members of monochronic cultures are more punctual, they separate private from work time, etc. Those in polychronic cultures are less punctual, and personal relations dominate the schedule. The mainstream cultures of the English language are much closer to the monochronic end of the scale. The second of Hall’s classic patterns differentiates **LOW-CONTEXT CULTURES**, where one tells things directly, and **HIGH-CONTEXT CULTURES**, where much needs to be inferred from the context. Slavic cultures are low-context. If a Russian dislikes something, he/she will tell you about it. Japanese culture is high-context. When a Japanese dislikes something, he/she will convey the idea via subtle contextual clues rather than directly.

Hofstede and Hofstede established the following dimensions: **POWER DISTANCE** (the level of tolerance for unequal distribution of power in the society), **INDIVIDUALISM** (feeling independent) versus **COLLECTIVISM** (feeling interdependent), **MASCULINITY** (the use of force is endorsed), and **UNCERTAINTY AVOIDANCE** (society’s tolerance to ambiguity and uncertainty). Later, Geert Hofstede added two additional dimensions: **LONG-TERM ORIENTATION** (the idea that one should prepare for the future) and **INDULGENCE** (the idea that it is a good thing to enjoy life). If we take a look at the mainstream culture in the United States, it is strongly individualistic, very tolerant to power distance, somewhere in the middle of the masculinity scale, tolerant to uncertainty, that is, low on the uncertainty avoidance scale, in the middle of the long-term orientation scale, and indulgent.

Hall's classic pattern and Hofstede's dimensions alike are underlying factors of some lexical phenomena. For example, high-context cultures develop fine-grained linguistic means of expressing things indirectly. Similarly, languages of collectivist cultures are likely to maintain more elaborate systems of kinship terms for a longer period of time in their development. To add another example, indulgent cultures are likely to have more elaborate networks of adjectives describing fun and enjoyment, etc. For those of you who were wondering about my unsolicited advice at the beginning of this chapter (that you did not follow, given that you are reading this), that too has to do with my coming from a collectivist culture, where giving unsolicited advice is the national pastime.

The notion of the lexeme (which is, as mentioned, a technical term for what is known as a "word" in common parlance) is most elusive. Various criteria have been proposed to define a lexeme. First, there is independent use. For example, the word *word* can be an independent utterance, and the sound *d* cannot. Another criterion is whether it is separated from other words in a text through pauses, or blanks in writing systems that have blanks. So, to provide an example, *this word*, not **thisword* (* means 'impossible'). Then, there is pronunciation unity, the fact that the sounds belonging to one word are pronounced or written together, without a break or something inserted (e.g., *word*, not **wo rd* or **woexrd*). Another criterion is the ability to change the meaning of other words (for example, *word count*, changes the meaning of *count*, and it does so differently than *number count*). Next, there is the ability to convey meaning. For example, the word *word* has a certain meaning; among other things it means a unit of the lexical system of a language, different from, say, the word *number*, which is a unit of the numerical system. Finally, its parts cannot be reshuffled, so, it has to be *word*, not **rdwo* or something like that. Various authors have ascribed varied degrees of importance to all those criteria. While they are all useful in differentiating lexemes from other linguistic elements, there is no single criterion or cluster that can unequivocally define a lexeme. This concept remains a prototype with a core of typical lexemes and a periphery where lexemes and non-lexemes (e.g., prefixes such as *bio-* and suffixes such as *-logy* in *biology*) overlap. One can assume various strategies in this field, but, for all practical purposes, I will consider lexemes to be those items one would normally look up in a dictionary: words, idioms, and lexical affixes,

such as the aforementioned *bio-* and *-logy* (but not grammatical affixes, such as *-s* in *he works*; syntactic frames, such as Subject Verb Object; collocations, such as *drink beer*, etc.).

As noted, vocabularies belong to their respective languages and cultures, and the need to talk about separate languages comes from the fact that there is a practical need to translate one language into the other and to learn languages different from one's mother tongue. Establishing links between any two out of some 7,000 world languages is a highly intricate task. Words are involved in the process primarily in establishing LEXICAL EQUIVALENCE. The essence of the process is to find the most appropriate lexeme in the TARGET LANGUAGE for the one in the SOURCE LANGUAGE. For example, when translating from Estonian into English, Estonian is the source language and English is the target language. In the process of establishing lexical equivalence, the situations where one would encounter one-to-one equivalence are extremely rare. Normally, something is always off (a word exists in one language and it is absent from another, a word in one language has more equivalents in another language, etc.), which makes translation and study of languages a difficult process.

These differences are a consequence of a complex organization of words. On the one hand, they feature complex internal semantic structure. Most of them are POLYSEMOUS, that is, they feature multiple senses. On the other hand, words are included in various external networks: they show affinity to combine with some words more than others, they share some components with other words, they are likely to be found in certain types of texts more often than in others, etc. The internal workings of words will be discussed in part B, and their relations with other words and contexts will be addressed in part C. Needless to say, the internal affairs of words are connected to their external affairs. For example, very often one sense of the word may show combinatory affinity of its own, as in *cool* being neutral when meaning 'cold' and informal when meaning 'good'.

B The Internal Affairs of Words

Special monographs and introductory textbooks about LEXICAL SEMANTICS (meanings of words), and those about SEMANTICS in general (meaning of all language elements, words included), have been advancing various models to account for the internal affairs of words,

their meanings. The same is true about monographs looking into cross-cultural aspects of semantics.

Many of these intricate models are certainly important in solving various theoretical linguistic problems. However, for the purposes of this book, a simple model proposed by Zgusta for practical work in lexicography will suffice. In this model, meanings of words have three primary components, **DENOTATION**, that is, what they signify, **CONNOTATION**, that is, what kind of attitudes they invoke, and **APPLICATION DOMAIN**, that is, to what they can pertain. For example, *food*, *chow*, and *feed* all have the same denotation. They signify the solid matter used to sustain living organisms. However, the connotation of the word *chow* is quite different, given that it is informal, so the attitude toward using it in formal situations is negative. Similarly, the application domain of the third word is much narrower, given that it is used in conjunction with food for domestic animals only. The cross-linguistic differences in the lexicon, as will become clear in Parts II through V, are most commonly caused by differences among denotation, connotation, or the application domain.

One important contribution to linguistic diversity of our world comes from different developments of new senses. New senses of words are developed from existing ones. This is known as **SEMANTIC TRANSFER**. For example, human and animal leg were first, and leg of a table was a new sense developed from the existing ones via semantic transfer. The most common mechanism of semantic transfer is **METAPHOR**, in which the link between the existing and the new meaning based on similarities of some kind is established. There are various approaches to metaphorical transfer. The simplest of them all, the two-domain approach, will suffice in navigating this book. The model encompasses the **SOURCE DOMAIN** (e.g., the human and animal leg in the aforementioned example), the **TARGET DOMAIN** (the leg of the table), and the **LINK** between them (the leg of the table looks like and performs a similar function of supporting something, just like the human and animal leg). Needless to say, these links can be different in different cultures.

Senses of each word can be organized in different ways. In some, there will be a central initial sense from which various other senses develop. For example, the word *bed*, riverbed, bed of flowers, bed of a roadway, bed as a layer of rock, etc. all developed from the initial sense of the piece of furniture. This is known as the **RADIAL MODEL**, where