



Basics of Linguistics **for future teachers of English**

Silvia Pokrivčáková



2023



Basics of Linguistics **for future teachers of English**

Silvia Pokrivčáková



2023

Author: prof. PaedDr. Silvia Pokrivčáková, PhD., © 2023

Reviewers:

prof. PhDr. Gabriela Lojová, PhD. (Comenius University in Bratislava)

prof. PhDr. Milan Ferenčík, PhD. (University of Prešov)

Acknowledgement

The publication is the result of the VGS project *Extension of study support for 1st-year students as dropout prevention* (2022-23) funded by the Faculty of Education, the University of Trnava, Slovakia.

Cover: Stock Cake, AI-generated abstract geometric art,
https://stockcake.com/i/abstract-geometric-art_497512_246801

Publisher

Published in 2023 by Gaudeamus, the publishing house of the University of Hradec Králové as its 1,858th publication.



Univerzita Hradec Králové
Nakladatelství Gaudeamus

ISBN 978-80-7435-924-8

Copyright information

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the acts mentioned above, without the author's prior permission.

The author's right to be identified as the Author has been asserted following the Copyright Act No. 618/2003 on Copyright and Rights Related to Copyright (Slovakia, EU). The text is assumed to contain no proprietary material unprotected by patent or patent application; responsibility for technical content and protection of proprietary material rests solely with the author and is not the publisher's responsibility. It is the author's responsibility to obtain all necessary copyright release permissions to use any copyrighted materials in the manuscript before submission.

The electronic textbook ***Basics of Linguistics for future teachers of English*** (ISBN 978-80-7435-924-8) is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs Licence. Based on a work at <https://pdf.truni.sk/elektronicka-kniznica>. In case of specific questions regarding re-sharing the contents or their fragments, please get in touch with the authors directly.



The textbook is accompanied by the audiobook ***Basics of Linguistics for future teachers of English*** (ISBN 978-80-7435-924-8), licensed under a Creative Commons Attribution-NonCommercial-NoDerivs Licence. Based on the work at <https://pdf.truni.sk/elektronicka-kniznica>.

Table of contents

	Introduction	7
1	Communication	9
2	Language	16
3	Languages in the world	23
4	Linguistics	28
5	The sound of language	31
6	Language structure	38
7	Meaning in language	51
8	Writing systems	62
	Conclusion and future trajectories in linguistics	66
	References	68

Introduction

Many aspects of human life depend on successful communication, from ensuring survival to building and maintaining rich social relationships. For their communication, humans have a nearly perfect tool – language. Language enables people to cooperate, reach their personal or academic goals, develop a successful career, and find suitable spouses. Because of this, knowledge of language and suitable communicative proficiency can significantly impact the individual's well-being and success. The lack of verbal skills can cause serious problems in all the above-mentioned areas. This also determines the importance of the systematic study of language.

The textbook *Basics of Linguistics for future teachers of English* serves as a primary information source for the subject of the same name. It functions as the introduction to the study of language for students who study English-focused philological study programmes. Through the course, students are encouraged and supported to fulfil the following learning objectives:

- to gain basic knowledge of linguistics and the latest research findings in the field of linguistics,
- to learn about fundamental relationships between linguistic disciplines,
- to understand basic linguistic terminology,
- to be ready for further study of the English language in specialised courses on English phonetics and phonology, morphology, syntax, lexicology, and other linguistic disciplines,
- to develop crucial academic skills necessary for independent study.

The content of the textbook is divided into 8 chapters. The first chapter positions the study of language in a broader academic perspective of the theory of communication and semiotics. It shows how language relates to other modes of human and non-human communication. The second chapter introduces basic myths and research theories explaining the origin of languages, continues with a brief overview of contemporary definitions of language, and concludes by outlining constitutive features which distinguish language from other means of communication. The third chapter briefly overviews languages worldwide, language families and groups, their classification and typology. The fourth chapter deals with linguistics as a scientific branch that studies various language aspects through research methods. The equal attention is paid to the branches of general and applied linguistics. In the following chapters, the sound of language (Chapter 5), language structure (Chapter 6), and language meaning (Chapter 7) are discussed. The chapters synthesise the latest knowledge from respective disciplines

of general linguistics (phonetics and phonology, morphology, syntax, and lexicology) and some disciplines of applied linguistics, mainly anthropological and cultural linguistics. The eighth chapter focuses on written language as a younger form of verbal communication and introduces various writing systems that have developed during the history of humankind.

For better orientation and to support students' independent study, each chapter is preceded by the explicitly expressed study objectives, areas of study (or linguistic branches), and key terms related to the respective chapter.

The author would like to express her sincere gratitude to all those students and colleagues who inspired her to teach the course and write this book. A special thanks goes to doc. Mgr. Hana Vančová, PhD., doc. PhDr. Božena Horváthová, PhD., doc. PaedDr. Rastislav Metruk, PhD. and doc. PaedDr. Zuzana Hrdličková, PhD. for their careful reading and expertise comments from the perspective of linguistic courses teachers. The author also cordially thanks the reviewers, namely Prof. PhDr. Gabriela Lojová, PhD. from Comenius University in Bratislava and Prof. PhDr. Milan Ferencík, PhD. from the University of Prešov, for their valuable and constructive notes, which helped to improve the text's quality.

Author

1 Communication

Study objective(s):

This chapter aims to position language study in a broader academic perspective of the theory of communication and semiotics. It shows how language relates to other modes of human and non-human communication.

Areas of study: theory of communication, semiotics

Key terms: communication, non-human communication, human communication, types of human communication, non-verbal communication, verbal communication, signal, sign, symbol, natural sign, arbitrary sign

Communication

A developed communication system is one of the most critical conditions for a successful biological species. Communication is also an umbrella term and the most general frame for any discussion about language. That is why this textbook begins by defining this term and concept.

In the most general sense, **communication** is any transmission or exchange of information (meaning) between individuals (humans, members of other species, or non-living entities such as machines) through natural or agreed (arbitrary) means. **The theory of communication** (which is a scientific discipline focusing on the matter of communication) distinguishes between human and non-human communication.

Non-human communication

Non-human communication includes plant communication, animal communication, and machine (digital) communication.

Plants can communicate using volatile organic compounds, electrical signalling, standard mycorrhizal networks, or sounds. Experts believe plants can share information about available nutrition sources and dangers which may start coordinated group behaviours (for more, see Gagliano, 2013; García-Servín et al., 2021; Heil, 2009; Sakurai & Ishizaki, 2024; Yang, 2003).

Animal communication happens when one animal transmits information to another, causing some change in the animal's behaviour that gets the information. Animal signals include *visual cues* (movements, gestures, facial expressions, body postures, and colouration, e.g. dancing bees, colouring during mating rituals of birds and frogs), *chemical cues* such as pheromones (e.g. dogs marking their territory by peeing on bushes and lampposts; ants following an "invisible" trails leading to food); *auditory cues* (e.g. chirping birds, barking dogs), *tactile cues* (through stroking rubbing, touching, vibrating, and pressure, e.g. see monkeys delousing each other), or *electric cues* (e.g. some fish).

Animals communicate to learn about their surroundings, obtain food, find mates, establish dominance or defend their territory, share information about danger, coordinate group behaviours, care for young, etc. In general, the primary purpose of both plant and animal communication is to carry out one's life functions and ensure the species' survival.

Although much has already been learned about nonhuman communication, much more remains to be discovered. The research area is increasingly studied since it is generally believed that understanding this type of communication will support a better understanding of how humans communicate (Deverell, 2002).

While the two previously discussed types of non-human communication have been naturally evolving for millions of years, the last non-human kind of communication is entirely artificial, human-made. In **machine communication**, electric signals are created by interruptions in a constant energy transfer. Communication channels are either analogue or digital. Machines are embedded with sensors, software, and other technologies to communicate. Today, everyday objects and tools with sensors not customarily considered computers are connected to networks, usually through the cloud-based internet, allowing them to generate, exchange and consume data with minimal human intervention (Pretz, 2013; Rose, Eldridge, & Chapin, 2015). These networks create "the Internet of Things" (IoT). "Within an IoT, all things are able to exchange data and if needed, process data according to predefined schemes" (Li, Xu, & Zhao, 2015).

Human communication

People can opt for numerous methods to convey or share information with other humans, e.g. visuals, personal appearance, haptics, proxemics, chronemics, gestures, movements, auditory means, or olfactory, electromagnetic and biochemical signals. Similarly to other biological species (plants and animals), humans use various means to interact and ensure the protection and survival of species through communication. They include kinesics (expressed by bodily movements, incl. gestures, postures, walking styles, and dance), proxemics (expressed by personal space), haptics (which refers to touching behaviour, like handshakes, holding hands, kissing, or slapping), paralanguage (a manner of articulation, lips control, rhythm, intensity, pitch, fluency, and loudness), chronemics (the specific use of time like moving slowly or quickly), and physical appearance. Artists and their audiences communicate with colours (pictures) and shapes (sculptures). Humans can also communicate with sounds like music and sound signals (alarms, sirens, etc.). All these are examples of **non-verbal** (or non-linguistic) communication.

However, compared to plants and animals, humans can exchange more complex ideas: beliefs, opinions, knowledge, wishes, commands, thanks, promises, declarations, or threats. They do so mainly via an extraordinarily complex communication system called language consisting of words as **verbal signs**. This type of communication is

called verbal (or linguistic) communication. **Verbal communication** encompasses linguistic communication through speech, writing, or gestures in sign language (Chandler & Munday, 2011).

According to the locus of activity and the number of people involved, four main types of human communication can be recognised (Oyeleye, 2004):

- a) intrapersonal – when communication takes place within a person,
- b) interpersonal – when communication passes between two or more people,
- c) group – when information is shared within a small group of people,
- d) mass communication – means sharing information within a large group of people.

In recent decades, a fascinating issue of possible interstellar communication with extraterrestrial intelligence has been discussed (cf. Vakoch, 2014).

Signals, signs, and symbols in human communication

As Sittig (2017, p. 35) has it, humans use “mutually agreed upon words, sounds, pictures, gestures, or behaviours to convey an intended meaning (e.g., thoughts, feelings, findings, or ideas) from one group to another” and unlike non-human communication, human communication is “open for numerous interpretations due to its extensive use of abstract language constructs involving words, signs, symbols, or sounds”. Similarly, Higgins and Semin (2000, p. 2296) state that humans “exchange information and influence one another through a common system of symbols and signs.” In both definitions, signs and signals as means of human communication are emphasised; therefore, it is time to explain their differences.

Signals are fundamental for all types of human and non-human communication, see the previous examples about plant and animal communication). However, signals, as the simplest of the three, can convey only simple messages. Signals can be the results of changes in fundamental physical characteristics (e.g. colour, size, shape, setting, or location, like in the case of flags and traffic signs) or by the change of a single environmental factor, e.g. presence versus absence of the light, sound, smoke, chemicals, etc. Human communication thus can include alternation of light and darkness (light signals), taps on wood (sound signals), puffs of smoke (smoke signals), and scents when signalling the unwanted leaks of cooking gas in the kitchen.

The sign may be any object, event, sound, shape, colour, or other entity which stands for something else. Compared to signals, signs can bear more significant amounts of meaning. The most common signs people encounter daily are pictures (including drawings, pictograms, logos, and other visuals), some human postures (e.g. a clenched fist as a sign of anger), and words.

The theory of communication distinguishes two main types of signs: natural (primary) and arbitrary (conventional) (see Fig.1).

Natural signs refer directly to the external reality; e.g. footprints in the snow mean that somebody or something walked over here; heavy clouds are signs of future rain, or tree rings directly indicate the tree’s age.

Arbitrary (conventional) signs, on the other hand, refer to the reality indirectly, e.g. a white flag is an internationally agreed sign of surrender, even though there is no physical or logical link between an act of surrender, white colour or a flag as an object.

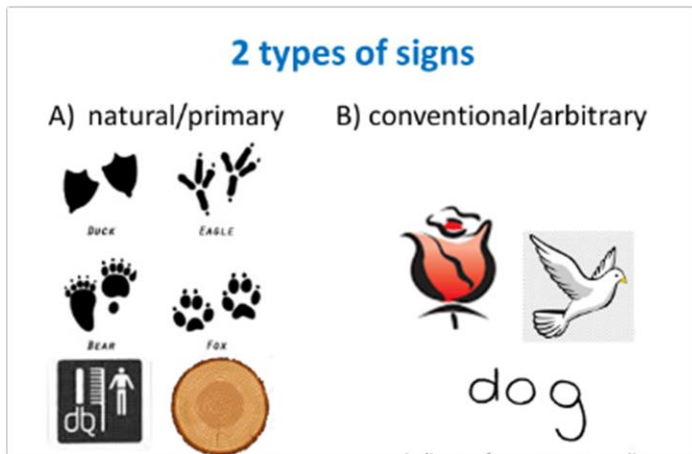


Figure1. Two types of signs in human communication

Similarly, a group of people decided that a friendly mammal used for guarding them would be called a dog. But another group of people decided it would be called “pes”, another “ein Hund” and another “Cachorro”. None of these words resembles the animal directly (compare Figure 2).

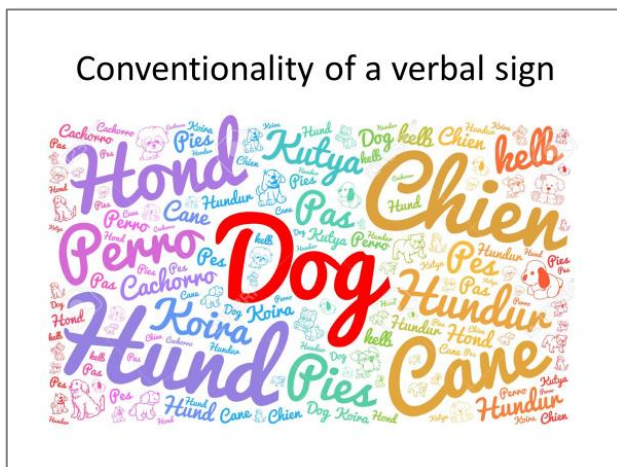


Figure 2.
Arbitrariness
(conventionality) of
the verbal sign
(source:
Shutterstock)

The previous example illustrates the arbitrariness (conventionality) of words. The wide array of words naming the same object led to the conclusion that words had been created within groups of people (tribes, nations), and their meanings were established by their habitual usage based on the mutual agreement of language users within those groups.

Signs are more complex and more challenging for abstract comprehension than signals. While many animal species can respond to signals such as smells, light blinking, and specific noises (whistles), only the most intelligent and trained animals (usually dogs, dolphins, and primates) can comprehend and respond to simple signs, e.g. whistling, gestures, etc.).

The meanings of signs can vary depending on the context or system they belong to. The signs “0” and “1” can serve as excellent examples. Both can represent numbers in a counting system (counting no item versus a single item). Or, they can refer to electrical signals in the binary digital system where “0” stands for “off” (no electric power inside the CPU) and a sign “1” stands for “on” (electric power in the CPU).

It is important to note that the meaning of a sign always depends on the existence of a shared coding system; e.g. for humans in some cultures, the sign “00” means a toilet. On the contrary, “00” always means just two zeroes (two interruptions) for machines.

In human language, all other types of signs (visual, sound, tactile, etc.) can be replaced by words as verbal (linguistic) signs. For instance, people who can speak English can use several words to name the sign “0”, e.g. “zero”, and usually the word “one” for the sign “1” (speakers of Slovak would use the words “nula” and “jeden”; German speakers would say “null” and “eins”, etc.). However, these words can carry other meanings in addition to the naming of arithmetical values. That is why human words take on the validity of symbols.

From the trio signals-signs-symbols, **symbols** are the most complex and challenging to understand. The symbol is “any device with which an abstraction can be made” (Akinola, 2023, p. 43). Symbols usually bear more than one possible meaning, e.g. the picture of a red rose represents a flower, but for many people, it also points to passionate love. Similarly, the image of a white dove can represent a bird, but in some contexts, it symbolises peace. Since words (as signs) typically have multiple meanings (denotative and connotative), they are generally considered symbols. Thus, when using words, humans communicate through **verbal symbols**. For instance, elaborating on examples from the previous paragraphs, the words “zero” and “one” do not only refer to numbers 0 and 1. The word “zero” in its symbolic meaning can also refer to a loser or a socially insignificant person. On the contrary, the word “one” can also mark the best, most important, or the most famous person in a lot.

Verbal communication

Verbal communication is the most effective form of communication, enabling humans to share the most complex thoughts, ideas, emotions, and other types of

information. Even though words are the basic building blocks of verbal communication, they constitute just one of its components. Other components include pronunciation (sound of language), morphology and syntax (structure of words and sentences, i.e. structure of language), and pragmatics (covering various conditions and situations in which language is used).

To express the nature of verbal communication in a simplified and more approachable manner, experts have been trying to design models of communication by outlining the main components of communication and types of interactions.

Shannon and Weaver (1948) proposed one of the most productive models, valued for its simplicity, clarity and generality. It involves five elements arranged in a linear order: a source, a message, a channel, and a receiver. The model is linear because the flow of information only goes in a single direction from a sender to a receiver without any feedback indicated (see Fig. 3).

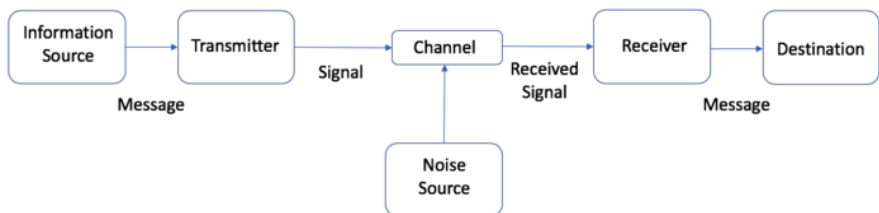


Figure 3. The linear communication model (Shannon and Weaver, 1949, cit in Appel-Meulenbroek, 2013)

Later, more elaborate communication models (interactive and transactional) were proposed. They include the same set of components; however, a feedback element and relevant contexts are added (compare Fig. 4). They express the idea that information (or “a message”) is transferred from a speaker (or “a source”) to a hearer (or “a receiver”) through a communication channel. A source and receiver must use the same code to convey a message effectively. A speaker encodes the message, and the receiver decodes it to understand it. At the same time, a hearer or hearer provides a speaker with continual feedback, usually offered through visual signs like eye contact, smiling, nodding their heads, etc.), or verbally by asking for clarification. Feedback helps immensely the effectiveness of communication; if missing, it is challenging for the speaker to guess whether the audience has the message and is happy with it. If the non-verbal message from the recipient is unsatisfying or ambivalent, the speaker can adapt their communication in the middle of sending a verbal message in response.

The transaction models of communication involve various internal and external conditions affecting the process and effectivity of verbal communication, e.g. physical, psychological, relational, social and cultural contexts).

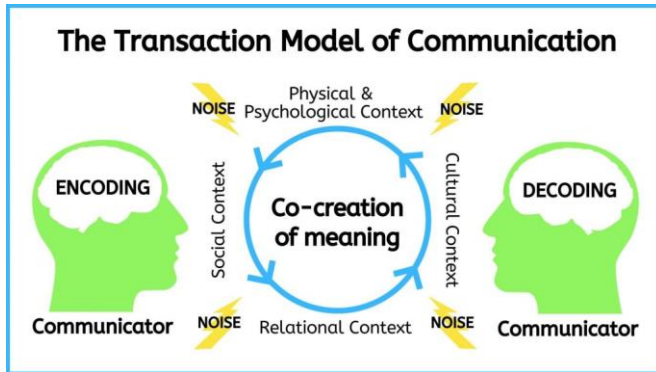


Figure 4. A scheme of a transactional model of communication (source: Lapum et al., 2020).

The interactive and transactional models are more realistic in expressing the dynamic nature of communication and its bi-directionality (when two communicators are involved) or multi-directionality (when more than two communicators share information). The overview of the main differences between linear, interactive and transactional communication models is represented in Table 1.

Aspect	Linear model	Interactive model	Transactional model
Nature of communication	One-way	Two-way	Simultaneous and continuous
Feedback	No feedback	Feedback is exchanged	Feedback is exchanged continuously
Role of participants	Sender → Receiver	Sender ↔ Receiver	Sender ↔ Receiver simultaneously
Message flow	Unidirectional	Bidirectional	Multidirectional
Contextual factors	Minimal emphasis on context	Acknowledges context in communication Strong emphasis on context (social, relational,	Multidirectional
Complexity	Simple and straightforward	More dynamic and responsive	Complex and multifaceted
Use cases	Announcements, PR, Marketing	Conversations, meetings, customer service	Negotiations, collaborative projects, real-time feedback

Table 1. Differences between the three models of communication (source: Sarkar, 2024)

2 Language

Study objective(s): In this chapter, language is discussed as the foremost and most specific method of human communication. First, some examples of myths and main scientific theories explaining the origin of language are introduced. Then, an overview of contemporary definitions of language is provided. Finally, fundamental language characteristics are compared to non-human (animal) communication features.

Linguistic disciplines: general linguistics, anthropological linguistics

Key terms: language, oral language, origin of language, protolanguage, full language

In the broadest sense, language is any complex system of signs used for communication. The theory of communication distinguishes several types of languages, while the type of a language is determined by the kind of signs in the system, e.g.:

- phonemes and words constitute *a spoken language*;
- mathematical signs construct *a mathematical language*;
- binary code made of digits 0 and 1 constitutes *computer language*;
- an intentionally created system of whistles establishes *whistled language* (such as La Gomera in the Canaries)
- gestures are part of *sign language*, etc.

At the heart of our concern lies a **spoken language**, which is a mode of communication based upon words and the combination of words into sentences (the term “spoken language” originates in the fact that all natural languages are primarily spoken and to this day many of them have no written form).

2.1 Origin of language

Although all species communicate, human communication is outstanding for its richness, flexibility, precision, and, most importantly – for” humans’ unique ability to use language” (Krauss, 2000). For millennia, people were contemplating about where their languages came from. Their ideas, beliefs and often very creative solutions were transformed into numerous **myths** (a myth = an ancient story explaining the origin of something).

Some myths claim the existence of human language right from the beginning of human civilization, others see languages as gifts (or curses) given to humans by gods, and others say that human skills to communicate in language(s) occurred due to some miraculous events.

The Bible tells the story of God asking Adam to name the creatures God had created, thus originating human language that is common to all people. Much later, after the great flood, the people of Babylon in Mesopotamia decided to build an enormous tower to touch the sky. God, upset with their pride and arrogance, punished them by mixing their language. The men stopped their work because they could not understand each other and ran to all parts of the world.

Somehow, similar stories framed by the great flood are told by Aztecs in Central America and the Kaska people of North America. According to the Aztecs, only one man, Coxcox, and his wife, Xochiquetzal, survived after the great flood. They floated on a boat to a new land where they gave birth to many children. After a dove occurred, children started talking different languages and could not understand one another. The Kaska people believed that there was only one country before the flood. People lived there together in harmony, and they shared only one language. However, after the flood, the people drifted worldwide, resulting in the birth of new tribes and many new languages.

Bantu people in East Africa tell the story of a severe famine that their people suffered long ago. Before that, all the people spoke only one language. But during the flood, people went mad, and the insanity made them run in all directions. They started babbling in different languages.

One tribe of Aborigines in Australia preserve a rather appalling myth about a malicious old woman called Wurruri. When she died, people were celebrating. During celebrations, they ate her body and started talking in various languages, according to which part they ate.

In Hindu mythology, the origin of language goes to the tree of knowledge, which grows in the centre of the earth. It was magnificently tall and reached nearly to heaven. The proud tree believed it should gather all men under its shadow and protect them. God Brahma punished the proud tree by cutting off its branches and throwing them down on the earth. The branches grew into banyan trees, separating men into many groups with different beliefs, customs, and languages.

Ancient Greek mythology states that from the beginning of humankind, people lived under the rule of Zeus and spoke only one language, which was given to them by God Philarios and Goddess Philarion. The unity and harmony among people were lost when God Hermes brought them many languages.

In Norse mythology, Bor and Bestla were the parents of three prominent gods: Odin, Vili, and Vé. While on a journey, they came across three trees. They decided to turn them into men. Odin gifted them spirit and life; Vili gave them wit and feeling, and Vé conferred them with speech, hearing, and sight.

This is just a brief sample of myths explaining the origin of language (for more, study Ganpule, 2013). While they are different, some share a similar motive of an initial harmony disrupted by some magical event. People started talking in many languages, leading to the original peace and general understanding.

Even though myths offer supernatural explanations for the origin of language, their study is not scientific without any value. As Barnard (2010, p. 17) has it, mythology does not perhaps explain the origin of language, but it does partly explain why language needs to be as complex as it is”.

Theories of language origin

Numerous thinkers, philosophers, and scientists have tried to explain the origin of human languages using empirical evidence, logical thinking, and research methods from history, philosophy, philology, anthropology, and human physiology. For instance, *comparative biology* studies what changes have occurred due to evolutionary pressures in the human lineage alone since no other living species has language. As a result, numerous scientific theories have been formulated; however, none have been proved irrevocably as perfectly correct and valid. The secret of the origin of language remains unsolved. In the following part, the most prominent theories are introduced, as summarised by Yule (2020, pp. 1-8).

1) The theory of divine source

Similarly to above-cited myths, advocates of this theory argue that the human ability to communicate either in oral or written form was given and protected by the gods who were associated with the concept of general wisdom, as well, e.g. *Saraswati*, the Hindu goddess of knowledge, creativity, and speech; *Nabu*, the Babylonian god of wisdom and writing; *Al-Kutbay*, the Nabataean god of knowledge and writing; or *Nisaba*, the Sumerian goddess of writing, learning, and the harvest, Egyptian *Seshat* was the goddess of wisdom, knowledge, and writing; Armenian *Tir* was the god of wisdom, written language, rhetoric, schooling, and the arts, etc. This group of theories include those who see language as a gift given to primitive humans by extraterrestrial visitors (e.g. Erich von Dönniken).

2) The theory of natural sound source

According to this theory, the first words originated as imitations of natural sounds (the onomatopoeic nature of a language). “In this scenario, when different objects flew by, making a caw-caw or coo-coo sound, the early human tried to imitate the sounds and then use them to refer to those objects even when they weren’t present” (Yule, 2020, p. 3). Some authors add the hypothesis that language was born from primitive vocal emotional responses to pain or joy, see so-called “ouch-theory”.

3) The theory of the musical sources

This theory is based on the belief that the first people charmed each other with musical and rhythmic sounds, which later evolved into words and sentences. Thus, the words’ primary meanings were coded into intonation and melody of speech. The phonemes developed much later by articulating the air stream leaving from the lungs.

4) *The theory of the social interaction source*

The theory reflects that the first language was developed when interacting with others in social groups. Members of the group started using various sounds with the agreed meaning, which also explains the arbitrariness of language.

5) *The theory of physical adaptation source*

This hypothesis focuses on the impact of the physiological transformation of the first people from four-legged animals to two-legged creatures. The change of posture and related physiological modifications led to the change in breathing, the change in morphology of vocal chords, and the change in how oral sounds were created.

6) *The theory of the tool-making source*

This theory links the birth of language with the ability of the first humans to make tools. Toolmaking led to the brain's further development and consequently became lateralised. Thus, a space for the ability to speak (Broca's area) evolved.

7) *The theory of the genetic source*

Representatives of this theory believe that language is genetically coded in a human being. It is proposed that human babies are born with a unique capacity for language. Language is considered an innate skill (no other biological species has it). This theory has been enforced lately by discovering specific alternations in the FOXP2 gene, which are thought to be essential for developing spoken language (Staes, Sherwood, Wright *et al.*, 2017).

Further development of language

Regardless of the lack of unity in opinions about the very first moments of language, most researchers agree that modern language was preceded by a prolonged period involving a more primitive pre-language (Bickerton 1990; Calvin and Bickerton 2000; Jackendoff 1999, 2002: Chapter 8; Wray 2002; Tallerman 2007). This has become known **as protolanguage**. They date proto-language to at least 500 thousand years ago (some authors go as far as 2 million years ago). Proto-language was probably structureless, consisting of individual proto-words or combinations that did not require full syntax (Bickerton, 1990; Jackendoff, 2002).

Full language gradually emerged from protolanguage, the tempo depending on the existing level of social structures and social intelligence. Jackendoff (1999, 2002) pointed out to various **linguistic fossils** – ancient features remaining in modern languages, such as 'defective' lexical items (ouch, hey, shh, dammit, etc.), which lack syntactic relationships. It has been agreed that morphemes were developed from phonological segments and phrases and sentences from words. The proper syntax could emerge only when the precursors to modern functional elements and grammatical affixes evolved - via the well-known processes of **grammaticalisation** (Bickerton 2000; Comrie and Kuteva 2005; Heine and Kuteva 2002, 2007; Jackendoff 2002). The oldest known full languages include Sanskrit, Sumerian, Hebrew, and Basque.

2.2 Defining language

Language can be and has been defined in many ways (which is a feature typical for humanities), and no definition is perfect. “There are numerous linguistic theories, each postulating and operating with its own definition of language, or its own perspective on language. At the same time, this presupposes that there are also many definitions of language that represent various linguistic movements” (Repka, 2020, p. 150).

Language is viewed in various theories as “a genetic inheritance, a mathematical system, a social fact, the expression of individual identity, the outcome of dialogic interaction, a social semiotic, the intuitions of native speakers, the sum of attested data, a collection of memorised chunks, a rule-governed discrete combinatory system, or electrical activation in a distributed network” (Cook & Seidlhofer, 1996, p. 4).

For example, one of the older definitions by Sapir (1921, p. 8) says that “language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntarily produced symbols.” Twenty years later, Bloch and Trager (1942, p. 5) claimed that a language was “a system of arbitrary vocal symbols by means of which a social group cooperates.” Twenty-five years later, Hall (1968, p. 158) defined language as “the institution whereby humans communicate and interact with each other by means of habitually used oral-auditory arbitrary symbols”. Contemporary *Oxford English Dictionary* (online) defines language as „the system of spoken or written communication used by a particular country, people, community, etc., typically consisting of words used within a regular grammatical and syntactic structure“. Gordon (2024) characterizes language as “a universal channel of communication into which various societies dip differentially to expedite and specify the numerous points of contact between individuals”.

Finally, Crystal and Robins in *Encyclopaedia Britannica* (2024) define language as “a system of conventional spoken, manual (signed), or written symbols by means of which human beings, as members of a social group and participants in its culture, express themselves. The functions of language include communication, the expression of identity, play, imaginative expression, and emotional release”.

After comparing the definitions mentioned above, their diversity and heterogeneity are apparent, but they share some key terms and concepts: human, communication, conventionality, arbitrariness, signs and symbols.

2.3 Human language properties

Even though all biological species can communicate, none are known to employ a communication system as complex as human language. The following list summarizes the main properties which distinguish human language from other types of communication (cf. Hockett & Altman, 1968; McGregor, 2009; Waciewicz & Żywicznyński, 2015; Yule, 2020):

- **Arbitrariness:** this property is related to the fact that there is no natural or inherent connection between the word’s form and its meaning. Verbal communication

requires cooperation and agreement among members of social groups. The only exceptions from arbitrariness are onomatopoeic words in which the connection between their sound form and meaning can be traced, i.e. their pronunciation resembles their meanings, such as in words knock, boom, crash, quack, splash, etc.

- **Displacement:** while animals' communication (via signals) is related to the immediate time and place where the communication occurs, humans can use language to refer to events remote in space and time. Thanks to language, people can discuss their past stories and future plans or make up stories and talk about events that never happened and likely will never happen. Moreover, language enables its users to talk nonsense or to lie.
- **Productivity:** humans continually create new words and utterances to name new objects or express new ideas. Due to the richness of verbal repertoire and openness of language as a system, the potential number of utterances humans can create is actually infinite. The productivity of language is closely related to its **discreteness** (language consists of isolatable, repeatable units such as phonemes, morphemes, words, and sentences that can be combined) and **duality of patterning** (forms of existing words can be recombined to form new meaningful units, e.g. spot or tops, or pots).
- **Reflexivity:** humans can use their language to talk about language itself, which enables them to think about the purpose, modes, and effectivity of verbal communication and thus further refine their verbal communicative skills.
- **Cultural transmission:** this property accounts for the fact that language is passed on from one generation to the next with encoded centuries- and millenia-long experiences, knowledge and culture.
- **learnability:** human languages can be learnt by various people, and simultaneously, humans can learn different languages (other creatures are limited to their genetically specified ways of communication)
- **specialization:** human language serves no other function but to communicate
- **interchangeability:** the roles of one language user can be changed anytime, i.e. a person can act as a speaker or a hearer and change these roles within one conversation.

Communication between humans and animals

Many people believe that primates can communicate very similarly to humans. Yet, primates are physically unable to produce human speech sounds. In the wild, they use only communicative gestures. For these reasons, modern studies and psycholinguistic experiments with non-human primates focus on teaching animals sign language or other visual communication systems. Examples of such studies are Washoe the chimp and Koko the gorilla experiments (Gallagher, 2014), during which both animals learned 100-200 signs. In the Nim Chimsky experiment, an animal subject could perform simple repetitions after the trainer (for more information, see The Gorilla Foundation website,

online). On the other side, many experts opposed such experiments and their conclusions. Most serious issues with primate studies included the complaints that the experimental primates needed extensive training, which, in addition, focused on particular templates. Grammar acquisition, such as sentence structures, was lacking altogether. Consequently, experimental results were anecdotal rather than experimental. Most animals merely imitated sounds made by humans and exhibited just stimulus-response behaviour.

In any case, animals could not learn and use grammatically complex structures, create novel utterances, and understand complex concepts (such as bacteria, atoms, and space). None of the “educated” primates attempted to pass down their ability to sign to their young, and thus, claims that primates had learned or acquired human language were greatly exaggerated.

However, other experiments proved that some primates could learn sign language and associate meaning with signs/words. For example, Rico the border collie could recognise 200 words, and Kanzi the bonobo learned 3,000 lexigrams (verbal symbols).

3 Languages in the world

Study objective(s): The chapter discusses the multiplicity of human languages and related problems of linguistic identification and typology. The process of language evolution is illustrated by the difference between live, endangered, dead, and extinct languages. The chapter also outlines the frame classification of world languages into families. Finally, the problem of language variants and dialects is indicated, too.

Linguistic disciplines: historical linguistics, comparative linguistics, anthropological linguistics, language typology

Key terms: natural language, artificial language, linguistic universals, language ecology, language evolution, living language, dead language, extinct language, endangered language, language family, language subfamily, language isolate, language variant, dialect

Linguistic diversity

In previous chapters, the main properties of human language were introduced. Based on a comparison with animal communication, another property can be observed – the diversity of languages. While dogs worldwide can bark in the same manner and understand each other regardless of the continent, country, or region they come from, humans have developed a wide array of languages, and it is likely that nobody can communicate in all of them.

Based on the most recent estimates and official statistics (UNESCO World Atlas of Languages, online), there are around 8324 languages, spoken or signed, which have been documented by governments, public institutions and academic communities. Of these, around 7000 languages are still used and spoken globally. Similarly, *Ethnologue*, the encyclopaedia of languages of the world (Eberhard, Simons, & Fennig, 2024), recognizes and describes 7,106 languages as of 2014. However, the exact number remains uncertain due to the complexities of the methodology used to distinguish between distinct languages, their variants, and dialects. In addition, some undiscovered languages may still be hidden in remote parts of the world.

The task of identifying and classifying languages based on comparing their phonology, morphology, syntax and lexicons goes to **comparative linguistics**, a branch of historical linguistics that also seeks to establish the historical relatedness of languages. It studies similarities and differences among languages but also searches for features shared by all languages. Such features are marked as **linguistic universals**.

It is important to note that comparative linguistic statistics do not include about 200 **artificial languages** created since the 17th century, mainly by philosophers, linguists, or

writers (e.g. J. R. R. Tolkien invented 13 artificial languages as the means of communication in his Middle-Earth universe). The most commonly spoken artificial language is Esperanto, with speakers between 100,000 and two million (Wandel, 2015). It was created by L. L. Zamenhof in 1887 as a blend of Latin, Romance, English, and German.

Comparative linguistics studies also **language ecology** (how languages interact with each other and the places they are spoken in) and **language evolution** (how the world's languages have evolved over time and continue to evolve). Similarly to biological species, languages undergo continuous changes (Christiansen & Kirby, 2003; Johansson, 2005). The main factors contributing to the historical evolution of languages are movements of populations, language contacts, social and cultural evolution, and the influence of geographical surroundings.

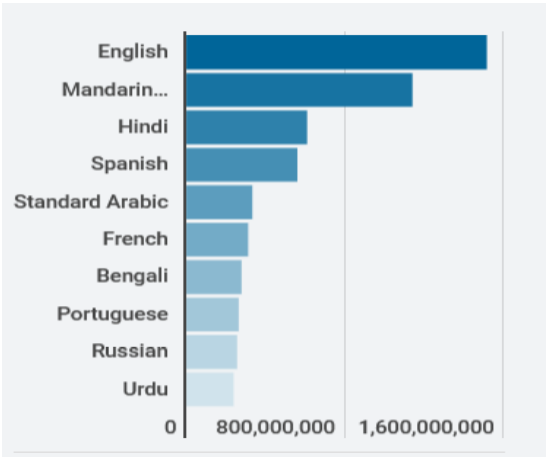
Some languages are more successful and used by more speakers than others. On the contrary, many languages are on the brink of extinction. Many projects are underway worldwide to prevent or slow the process of losing languages through intentional education; however, most endangered languages are unlikely to be saved or revitalised. Historical linguistics recognizes several stages leading towards language death.

According to the progress of their evolution stages, languages can have the status of:

- **living languages** which are currently used by a group of people as a primary means of their communication;
- some living languages are marked as **stable languages** because they are unlikely to disappear since all the children in the community of their users are still learning and using them;
- other living languages are **endangered**, i.e. languages at risk of disappearing because the young generation is not learning them and their older speakers die out;
- **dead languages** – which have no living native speakers but are still used in some specific situations, such as Latin in medical communication;
- **extinct languages** – currently spoken by no living native speakers and are not used in any communicative situation anymore.

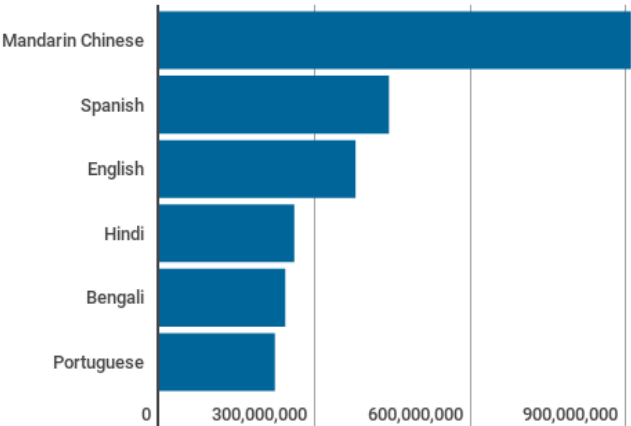
While the world's ten most spoken languages are used by nearly half of the global population (see Figure 5), most languages are spoken by fewer than 10,000 people (Austin & Sallabank, 2011). English has been the most spoken language worldwide for decades. Most spoken languages include Mandarin Chinese, Hindu, Spanish, and standard Arabic.

Figure 5. Top 10 most spoken languages in 2024 (source: Eberhard, Simons, & Fennig, 2024)



Not all speakers included in the previous statistics were also native speakers of those languages (a native speaker = someone who learnt a particular language as a baby and has been using it primarily in their everyday life since early childhood). The estimated numbers of native speakers of so-called “large” languages are indicated in Figure 6.

Figure 6. Languages with the most native speakers (source: Eberhard, Simons, & Fennig, 2024)



For languages, significantly uneven geographical distribution is typical. In 2014, over two thousand languages were used in Asia, from which about 800 were spoken in Papua New Guinea alone (see Table 2). The country with the world’s largest concentration of languages was the tiny state of Vanuatu in the Pacific Ocean (106 languages), while in Europe, only 285 languages were used.

Continent	Languages	
	number	%
Asia	2,303	32.4%
Africa	2,146	30.2%
Pacific	1,312	18.5%
North and South America	1,060	14.9%
Europe	285	4,0
Together	7,106	100

Table 2. Language distribution among continents in 2014 (source: Lewis et al., 2014)

Genetic classification of languages

Languages can be classified in many ways depending on the criteria employed. Historical and comparative linguistics work with **the genetic classification** of languages based on the fundamental hypothesis that languages evolved from **proto-language** (a common ancestor). It is believed that this first language was modified and changed within the transmission process from one generation to another. New modified languages may differ but remain genetically related, thus forming language families. They consist of a parental language (proto-language) and daughter languages. Languages on the same developmental level are called sister languages.

The Encyclopaedia of World Languages (Eberhard, Simons, & Fennig, 2024) lists 143 language families, six of them being the major ones:

- African / Niger-Congo (1,537 languages)
- Austronesian (1,225 languages)
- Trans–New Guinea (476 languages)
- Sino-Tibetan (457 languages)
- Indo-European (446 languages)
- Afro-Asiatic (377 languages).

They have the most speakers worldwide in different regions and countries.

Members of one language family must originate in the same proto-language, share the same phonological features, share similar vocabulary items (e.g. shared etymology), share grammatical features (e.g. flection), and share syntactical structures (e.g. word order). All five conditions must be met.

Even though most of the world's languages can be grouped into language families, a few have no known structural or historical relationship to any other language. They are known as **language isolates**. The examples are Basque (in northern Spain & southern France with 500,000 speakers), Burushaski (in north-west Kashmir, India & Pakistan with 20,000 speakers), Nahali (in India with less than 1,000 people), Ainu (in Hokkaido in Japan with just a few speakers), and Sandawe (in Tanzania with 60,000 speakers).

Language variants and dialects

One language has naturally several varieties characterised by their own phonological, syntactic, and lexical properties) but still comprehensible to other speakers. Language varieties include:

- **standard**, recognised by media, educational and cultural institutions as the desired mode of verbal communication
 - **sub-standard**, which usually exist only in a spoken form, include numerous deviations from standard varieties and are typical for informal registers of communication (e.g. dialects);
 - **regional**, associated with a specific place, country, or continent. For instance, English has ten **standard varieties related to** England, Scotland, Wales, Northern Ireland, the United States, Canada, the Republic of Ireland, South Africa, Australia, and New Zealand. Contemporary linguistics marks them as **different Englishes** (Crystal, 1999);
 - and **social**, related to a particular social group (e.g. dialects).
- Language varieties and dialects are studied by **sociolinguistics**.

4 Linguistics

Study objective(s): Language as a highly complicated human construct can be studied from many different points of view or at different levels, each contributing to a better understanding. The chapter briefly outlines linguistics as a science studying language and its various branches.

Linguistic disciplines: general linguistics, applied linguistics, sociolinguistics, psycholinguistics, neurolinguistics, cultural linguistics, anthropological linguistics, computational linguistics, forensic linguistics, translation studies, corpus linguistics

Key terms: grammar, stylistics, comparative linguistics, historical linguistics, psycholinguistics

In the previous chapter, the observant reader could notice that the term “*language*” was used with two different meanings:

- language as a universal means of human verbal communication (as discussed in Chapter 1),
- and language as a concrete manifestation of that universal capability within a specific social group (a tribe or nation), such as English, Slovak, German and more than 8,000 other languages (see Chapter 3).

Language (in both senses) as a highly complicated human construct can be studied from many different points of view or at various levels, each contributing to a better understanding. The scientific study of language appertains to **linguistics**. It aims to observe, analyse, and understand the units and patterns within and across languages.

Linguistics is a relatively young discipline. Before the 19th century, philosophers and philologues studied language as an abstract idea. **Philosophy** questioned the suitability of language as the only tool for expressing general questions about existence, knowledge, values, and life. As part of classical and literary scholarship, **philology** contemplated language using critical thinking and logical operations.

Linguistics, as a systematic study of language based on objective research and empirical data, was born in the 19th century. Since then, linguistics witnessed the rise of three major schools of linguistics. *The structuralist school*, represented by de Saussure (1916) and Bloomfield (1933), promoted the study of language as a self-contained structure that is independent of other aspects of human existence and verbal communication. Special attention was paid to observing and describing relationships between the elements within the structure. *The functionalist school*, developed by Halliday (2013) and Searle (1969, 1992, 1999), studied language through the prism of its

primary function as a communication tool. The approach focuses on the impact of social context, usage, and the communicative functions of language, such as cognition (relating information), expression (indicating mood), and conation (affecting or changing the decisions, actions, or beliefs of others through persuasion, advocacy, or pressure).

The generativist school of linguistics, represented by Chomsky (1965, 1993), search for and study the different phenomena that occur in all natural languages. Their results led them to suggest that all humans are capable of learning language because all languages are made up of certain shared (universal) rules. This led to the theory of “universal grammar”.

If linguists focus on general components, features, and characteristics of different languages, they devote themselves to **general linguistics**. Within it, they study the nature and fundamentals of languages, such as their origin, sound, structure (form), meaning, and usage. General linguistics includes descriptive linguistics and historical linguistics, too. Branches of general linguistics include:

- **phonetics** - the study of speech sounds in their physical aspects;
- **phonology** - the study of speech sounds in their cognitive/mental aspects;
- **morphology** - the study of the formation of words;
- **syntax** - the study of the formation of sentences;
- **lexicology** – the study of vocabulary, words and their meaning,
- **semantics** - the study of linguistic meaning;
- **pragmatics** - the study of language use;
- **discourse analysis** – the study of communicative acts.

Those linguists who study language in real-life situations and concentrate on applying the general descriptions and explanatory theories generated by general linguistics to various aspects of everyday human life develop **applied linguistics**. Branches of applied linguistics integrate linguistic research with other scientific disciplines depending on the aspect of life they focus on. They include:

Sociolinguistics conducts linguistic research about social communication contexts between and within social groups. Sociolinguists observe different linguistic features related to class, race, sex, etc. Its research domains include sociolects (= characteristics of the language of specific social groups, such as slang and jargon), social and regional dialects, language varieties, language dominance, language prestige, language imperialism, language equality, speech community, bilingualism, and code-switching.

Psycholinguistics deals with mental processes related to language production, reception, and acquisition. It is part of cognitive studies, focusing on mental mechanisms that make it possible for people to use language.

Neurolinguistics – concentrates on language-related neural structures and biological processes in the human brain, mainly using medical methods. Special attention is paid to speech deficiencies and defects, such as various types of aphasia.

Cultural linguistics (ethnolinguistics)– integrates linguistics and cultural studies. It explores cultural models, which are associated with the use of language (e.g. language of folklore, language of cultural communities, cultural jargon). It also examines how various features of human languages encode cultural conceptualisations, including cultural schemas, categories, and metaphors.

Anthropological linguistics uses linguistic techniques to analyse and learn about ancient cultures without written records. Anthropological methods are employed to determine the role of language in forming individuals and communities.

Computational linguistics analyses and explains linguistic phenomena using mathematics and informatics (NLP = natural language processing). It is also involved in designing various tools for human-machine communication based on automatic speech recognition and synthesis. Its research domains include AI-powered language tools, chatbots, intelligent writing assistants, computer-assisted language learning, and designing computer languages.

Forensic linguistics provides police officers, lawyers, judges and juries with evidence based on careful and systemic language analysis. Its objectives include interpretation of intended meaning in oral and written statements (e.g., confessions), authorship identification, voice identification, interpretation of expressed meaning in laws and legal writings, analysis of discourse in legal settings, the language of the law (e.g., plain language), analysis of courtroom language used by trial participants (i.e., judges, lawyers, and witnesses), trademark law, and interpretation and translation when more than one language must be used in a legal context.

Translation studies (Translatology) systematically research translation and interpreting processes and products. It uses the methods of comparative literature, computer science, history, linguistics, philology, philosophy, and semiotics (Ferenčík & Bednářová-Gibová, 2024). Its current research domain is machine translation.

Contrastive linguistics seeks to describe the differences and similarities between languages that can occur at every level of linguistic structure: speech sounds (phonology), written symbols (orthography), word formation (morphology), word meaning (lexicology), collocation (phraseology), sentence structure (syntax) and complete discourse (textology). It closely collaborates with **corpus linguistics**.

5 The sound of language

Study objective(s): The chapter introduces the topic of sound characteristics of human speech. It discusses the area of speech sounds and suprasegmental features of human speech, mainly from the perspective of language typology.

Linguistic disciplines: phonetics and phonology, linguistic typology

Key terms: phonation, articulation, speech sound, phoneme, consonant, vowel, approximant, tone, intonation, rhythm, transcription, IPA

As explained in Chapter 2, language primarily originated as a spoken means of communication. Only much later, writing systems were developed in some languages, and only a few members of privileged groups could read and write. Even today, many languages remain unwritten. This means that the fundamental characteristics of language are determined by the conditions and requirements of speech, not those by writing.

Even though the sound of language is studied from many perspectives, the study of spoken language is focused mainly on two areas. The first is knowledge of the physiological and physical nature of speaking and hearing, with a special focus on the workings of the vocal tract and individual organs of speech. This area of study is called **phonetics**. The second area is interested in mental processes in the human brain, which results in using and comprehending human speech. This area of the language sound is studied by both **phonology** and **psycholinguistics**.

Speech sounds

It is quite a paradox that speech, essential in human life, is produced mainly by organs with very different original functions related to digestion or breathing. Only much later in their physiological development was the function of speaking superimposed on them as a secondary function. Another paradox is that the primary material of speech is a biological waste product – exhaled air.

Speech sounds are the results of resonance created by various movements and configurations of articulatory organs. Each configuration and movement of the vocal tract creates corresponding differences in the exhaled air movements and vibrations. These comprise and transmit sound. The process by which the vocal folds produce certain sounds is called **phonation**, and the process of modelling the exhaled air stream is called **articulation**.

The human vocal tract can produce a great variety of sounds, but each natural language uses just a tiny part of the tremendous phonetic potential. Natural languages usually have only several dozen sound units to build words and utterances.

Speech sounds have been described and classified by phonetics from an articulatory viewpoint (regarding how they are produced) and from an acoustic perspective (by reference to the resulting sound waves, their frequencies, amplitudes, and so forth).

The most basic classification of speech sounds differentiates between vowels and consonants. **Vowels** are the sounds produced without a significant barrier in an articulatory tract. They typically occur as the essential centres of syllables. In many languages, the shortest possible words are made of just one vowel, as in the English word “awe” or the French word “eau”, pronounced as the single vowel sound /ɔː/. If an obstruction occurs in the vocal tract during articulation, e.g. if the tongue touches the teeth or the mouth is closed, the produced sound is a **consonant**. The sounds ‘in-between’ (‘incomplete consonants’), articulated with a narrowed but not closed-by-obstruction vocal tract, are called **approximants**, for example, the English sounds /w/ and /j/ (Ladefoged, 2005).

The speech sounds which distinguish one word from another are called phonemes. The number of phonemes worldwide is not fixed, as it varies by language. It is impossible to say how many phonemes there are in all world languages since the notion of **phonemes** only makes sense within a language (never across languages). For instance, sounds /l/ and /r/ are phonemes both in English (raw-law) and Slovak (lev-rev); however, they are not phonemes in Japanese because there is no minimally distinct pair of words (hence the research method of minimal pairs = observing the context where the particular sounds appear and identifying pairs of words which differ from each other only by one single sound).

Many people are curious about the number of speech sounds or the most frequent or rare sounds. The questions of this type are studied by **linguistic typology**, which seeks to classify languages based on identifying units and patterns in their structure or adherence to commonly accepted language trends and **universals** (items present universally in all languages). Interested readers should look up *The World Atlas of Language Structures* (WALS; Dryer & Haspelmath, 2013), a database of structural (phonological, grammatical, lexical) properties of nearly 500 languages gathered from published descriptive materials. Or they can opt for PHOIBLE, a cross-linguistic phonological inventory (Moran & McCloy, 2019). The UCLA Phonological Segment Inventory Database (or UPSID) is a statistical survey of the speech sound (phoneme) inventories in 451 of the world’s languages (Maddieson & Precoda, 1989).

So far, experts have not established the number of speech sounds humans use. The proposed numbers vary considerably from approximately 100 (as stated by *the International Phonetics Association*, which recognizes 87 consonants and 30 vowels plus a range of diacritics to add finer detail) to 800 (c.f. *canIPA Natural Phonetics* by Canepari, 2023). The situation is even more complicated with the attempts to establish the total number of phonemes. Even though all languages have phonemes, their total number cannot be stated since phonemes can only be defined within a concrete language as its contrastive sound categories. They cannot be compared or grouped across languages.

Therefore, only the numbers of phonemes in individual languages can be compared, and the numbers are astonishingly varied. The language with probably the least sounds is Hawaiian, which has 13 speech sounds, and Pirahã (one of the tribal languages used in Brazil), which manages with only 10 speech sounds: /a/, /i/, /o/, /p/, /t/, /ʔ/, /b~m/, /g~n/, /s/, and /h/.

However, the usual (i.e. statistically average) number of phonemes in individual natural languages is between 20 and 60. For instance, English has 44 phonemes (20 vowels and 24 consonants). Slovak uses the same number of phonemes but of different composition (15 vowels and 29 consonants; cf. Pavlík, 2004). The language with the most phonemes is often named Taa (also known as !Xuun or !Xóǝ), a Khoisan language spoken in Botswana and Namibia. It is estimated to have over 300 phonemes, including over a hundred types of clicks (Ladefoged & Maddieson, 2008).

Based on similar research studies, the **most common vowels** are /a/, /e/, /i/, /o/, and /u/. The most frequent vowel in UPSID is /i/, which can be heard in 393 languages and is closely followed by the vowel /a/, occurring in 392 languages (87%). Generally, the most common consonants across languages are /p/, /t/, /k/, /m/, and /n/. UPSID counted that the most frequently occurring consonant is /m/, which can be identified in 425 languages (94%). Then the sound /k/ follows (403 languages, 89 %).

Even though it is uncommon for a language to lack any of the five most frequent vowels and five consonants, none occurs in more than 7,000 languages worldwide. For instance, while standard Arabic is one of the “large” languages, it does not have /o/ and /p/, English does not have a single /a/, and Hawaiian has no /t/. Samoan lacks /t/ and /n/, while Quileute and Rotokas languages lack /m/ and /n/.

The following sounds belong among the **rarest vowels**:

- *voiceless vowels* like in Japanese (Ogasawara, 2005), or American Indian languages Zuni (Walker, 1972) or Cheyenne (cheyennelanguage.org),
- *aspirated, or “breathy”, vowels* like in Gujarati (Esposito et al., 2019) or Jalapa Mazatec languages (Kirk et al., 1993)
- *laryngealised, or “creaky”, vowels* present in Kedang, Coatzacoapan Mixtec and Otomanguean languages (c.f. Gerfen & Baker, 2005).

The list and geographical occurrence of **rarest consonants** (in a sample of 566 languages) can be consulted in Chapter 19A of the *WALS Online* (Maddieson, 2013), which identifies four classes of consonants that are missing in nearly 80% of the sampled languages (448):

- *clicks*, i.e. segments recognised by the IPA as [ǀ, ǁ, ǃ, ǂ], were studied by Brenzinger & Shah, 2023; Gil, 2013; Güldemann, 2007; Miller, 2011; Proctor et al., 2020 and others. They are present, for example, in isiXhosa, Kx’a, Tuu and Khoe-Kwadi language families in southern Africa (Bradfield, 2014), Chad (Lionnet, 2020), and Dahalo in Kenya;

- *labial-velar plosives*, e.g. [g͡b] and [k͡p]. They are pairs of consonants such as [g] and [b] or [k] and [p] pronounced simultaneously (Connell, 1994). They can be traced in the Niger-Congo and Nilo-Saharan language families of Africa, in the languages of Papua and New Guinea (Foley, 1986), in the Caucasian languages (Catford, 1977), and some Spanish dialects of Colombia, New Mexico, El Salvador, Ecuador, Chile, and Mexico (Mazzaro, 2010);
- *pharyngeals* are sounds pronounced in a pharynx, such as [ħ ʕ]. They are known to be present in Modern Standard Arabic, Somali, Iraqw, and the Nilo-Saharan language Tama (Shahin, 2011);
- and *dental or alveolar non-sibilant fricatives*, similar to the /ð/ sounds of English, occurring in Albanian, Icelandic, Modern Greek, and Spanish (Dediu et al., 2023).

Suprasegmental features of human speech

Phonemes are not the only speech elements that may influence the meaning. The same effect can be achieved using a melody used when pronouncing a syllable (i.e. **tone**) or a melody used when pronouncing a sentence (i.e. **intonation**). Languages which apply “lexical tones” (tones affecting the meaning of words) are called tonal languages, e.g. Cherokee, Mandarin Chinese, Thai, Vietnamese, etc. The frequently cited example is the changing pronunciation of four tones and related distinct meanings of the word “ma” in Mandarin Chinese (see also Figure 7):

- mā 媽 mother (level tone)
- má 麻’ numb (high-rising tone)
- mǎ 馬 horse (high-falling tone)
- mà 罵 to scold (low tone)

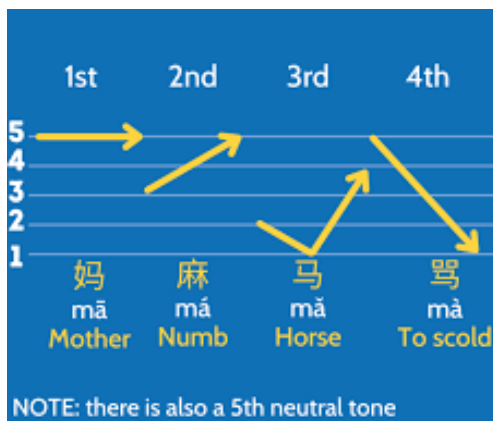


Figure 7. Tones in Mandarin Chinese
(source: Hobbs, 2023)

Intonation

In some tonal languages, e.g. in Ngiiti language, different tones can change the grammatical tense of sentences. Tonal languages also have intonation but are generally subordinated to the tones.

In most Indo-European languages, intonation does not change the lexical meaning of words, but it can modify the meaning of sentences. Intonation helps express the speaker's attitude or emotions (see the category of grammatical mood). The three main patterns of intonation are falling, rising, and fall-rise.

Language rhythm

In music, rhythm is associated with constantly repeating the same musical patterns. Language also has rhythm, even though it usually is not as regular or remarkable as musical rhythm (the only apparent exceptions may be poetry and song lyrics). Language rhythm depends on the sonority, length, and sequentiality of syllables.

According to the frequency and combinations of these features, languages can be divided into two categories: syllable-timed and stress-timed. While syllable-timed languages use a constant syllable length, stress-timed languages maintain a relatively constant interval between the stressed syllables.

As Goedemans and van der Hulst (2013) have it, modern linguistic theory recognizes **rhythm** as “a manifestation of the fact that syllables are further grouped into constituents called **feet**, which are usually binary groupings of syllables“. According to this aspect, languages can be divided into:

- **trochaic** – where a dominant foot is a trochee, which starts with a stressed syllable followed by an unstressed one. The pattern stresses every odd syllable from the left in languages with initial main stress. English, German, Slovak, Czech, and Hungarian are all trochaic languages.
- **iambic** – where iamb, a foot in which a stressed syllable follows the unstressed syllable, dominates. Languages with iambic rhythm typically stress even syllables from the left or odd syllables from the right. Iambic languages are represented by minor languages such as Seneca and Yupic.
- **dual** - combining the previous two systems, e. g. Wichita language;
- **undetermined** - with no clear dominant foot type, e.g. Irish, French, Maltese;
- and those with **no apparent rhythmic stress**, e.g. Catalan, Russian, and Turkish.

In their research on the sample of 323 languages, Goedemans and van der Hulst (2013) found that most languages prefer trochaic rhythm (compare Table 4).






Value	Representation
 Trochaic: left-hand syllable in the foot is strong	153
 Iambic: right-hand syllable in the foot is strong	31
 Dual: system has both trochaic and iambic feet	4
 Undetermined: no clear foot type	37
 Absent: no rhythmic stress	98
Total:	323

Table 4.
Rhythm types
(source: Goede
Mans & van der
Hulst, 2013)

Research of language sounds

As already mentioned, the sound of language is studied by the linguistic disciplines of phonetics and phonology. The research focuses on collecting, recording, documenting, and archiving sound samples of all languages (including their variants and dialects), as well as describing and analysing speech sounds and phonetic features of languages. Contemporary phonetics uses the latest Imaging and visualisation technology, such as digital oscillograms, spectrograms, and intonograms (compare Figure 8).

To document sounds precisely and make comparisons possible, sound recordings are translated into specific transcription systems that may differ significantly from standard orthographic writing systems (c.f. Pavlík, 2004).

The principal and oldest representative organisation for phoneticians and phonologists is the *International Phonetics Association (IPA)*. It was established in 1886 in Paris. The academic association aims to carry out and promote the scientific study of phonetics globally. It introduced the most respected universal phonetic transcription system, the *International Phonetic Alphabet* (also IPA, online).

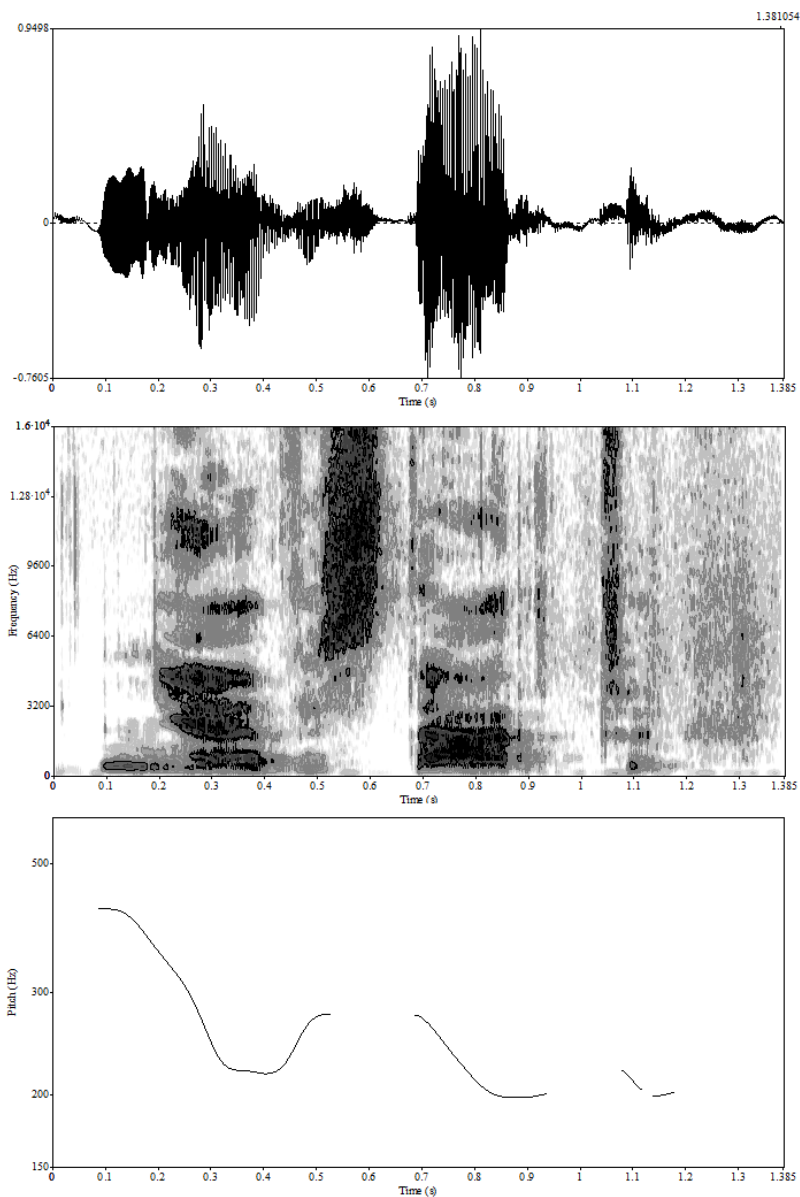


Figure 8. Oscillogram, spectrogram and intonogram for the same utterance: ‘nie, wystarczy’ /n’evystart^Sy/, which is Polish for ‘no, enough’ (source: Karpinski, 2014).

6 Language structure

Study objective(s): The chapter focuses on the ways and rules words, phrases, and other larger units are arranged for effective communication. The set of the said ways is called language structure, and the set of rules is marked as grammar. Much of the chapter introduces the main grammatical categories described from the language typology perspective. All language structures explained in this chapter are illustrated using examples from contemporary standard English.

Linguistic disciplines: general linguistics, morphology, syntax

Key terms: language structure, grammar, function words, grammatical meaning, morph, morpheme, allomorph, morphological processes, grammatical categories

To understand or produce meaning in any language, its user must know more than the words of that language (in the sense of both their pronunciation and lexical meanings). The practical language user also needs to master **language structure**, i.e. ways and rules of how individual words can be combined into larger units: phrases, clauses, sentences, paragraphs, and texts. Mastery of language structure helps speakers express their messages concisely and avoid unnecessary confusion and misinterpretation, which saves time and energy for both the speaker and the listener. The study of language structure is relevant for multiple subfields of linguistics, such as psycholinguistics, language acquisition, historical linguistics, and natural language processing in computational linguistics.

While strict adherence to language structure is not required and expected in private oral communication (possible deficiencies in verbal communication may be easily compensated by other means such as gestures, mimics, and visuals), it is crucial for professional communication, where clarity is vital.

To produce and understand language structure, the proficient language user needs to know various forms and functions of words in a particular language. This aspect of language is studied by the branch of linguistics called **morphology**. In addition, the effective speaker needs to know how to arrange words meaningfully into larger units. This aspect of language is explored by the linguistic branch named **syntax**. The rules that govern how words are formed, combined, and arranged in larger language structures in a particular language are called **grammar** (e.g. English grammar, Slovak grammar, etc.). In linguistics, the term grammar is understood in more ways. As mentioned in Britannica (2024), it names “rules of a language governing the sounds, words, sentences, and other elements, as well as their combination and interpretation. The word grammar also denotes the study of these abstract features or a book presenting these rules. In a

restricted sense, the term refers only to studying sentence and word structure (syntax and morphology), excluding vocabulary and pronunciation”.

Morphology

Generally, the **grammatical meaning of words** can be expressed in two ways:

- by separate **function words**, i.e. small words that are used to express some grammatical categories, for example, case, plural or definiteness (as “van” in Dutch, “de” in French, or “of” in English);
- or by changing content word forms (e.g. inflection and modification). These changes in word forms for grammatical reasons (to meet the grammar rules of a given language, e.g. agreement between a noun and a verb) are manifested by the changes in their specific characteristics called **grammatical categories**.

The inner structure of words (word forms)

The smallest unit of a word that carries a grammatical meaning is called a **morpheme**. For example, the word *international* contains three morphemes (i.e. elements with grammatical meaning): the prefix *inter-* (its grammatical meaning is “added meaning of between”), the root *nation*, and the suffix *-al* (the grammatical meaning of “an adjective created from a noun”). Each word contains at least one morpheme.

The following types of morphemes can be distinguished in the majority of languages:

- **a free morpheme** – can stand alone as a word (eat, water, free)
- **a bound morpheme** – is not a complete word on its own but does affect the meaning or grammatical category (e.g. English -s, -ed, -ing)
- **a root** - a morpheme bearing a lexical meaning which cannot be further divided into smaller parts (in-clude-s, inter-nation-al)
- **an affix** - a cover term for suffix, prefix, etc.
- **a prefix** - a morph that precedes a root (in-clude-s, inter-nation-al)
- **a suffix** - a morph that follows a root (in-clude-s, inter-nation-al)
- **a stem** - the part of a word responsible for its lexical meaning which remains unmodified during inflection, e.g. *insider*-s.
- **derivational morpheme** - changes the word class of a word (sad+ness, re+action)
- **inflectional morpheme** – the last morpheme in the word structure that modifies the word's grammar feature(s) but not its word class (cat+s, write+s).

The concrete spoken or written form of a morpheme is called a **morph**. Thus, it would be more appropriate to say the word “international” contains three morphemes expressed by the three morphs: inter + nation + al.

If a morph has multiple forms, they are called **allomorphs**, e.g. in written English, the prefixes “in” and “im” in English words *incredible* and *impossible* are allomorphs of one morpheme with the grammatical meaning of negation. On the other hand, the prefix “in”

in words *inability* and *insider* represents two morphs because they express two different grammatical meanings (negation versus inclusion).

Morphological processes

In various ways, words are formed and structured in a specific language through the addition, modification, or removal of morphemes are known as **morphological processes**. In other words, a morphological process is “a means of changing a stem to adjust its meaning to fit its syntactic and communicational context” (Matthews, 1991, p. 125). Morphological processes either alter the meanings (both lexical and grammatical) of existing words or help create entirely new words. Studying morphological processes leads to understanding how languages evolved and how they are used within linguistic communities. To learn more about the research into word-formation typology and tendencies, study Štekauer et al. (2012).

The most productive morphological processes in most languages are:

- **derivation**, i.e. adding affixes (prefixes or suffixes or both) to a root (a base word) while a derivate belongs to a different word class (part of speech), for example, from a noun “nation” a new adjective “national” was created by a suffix “-al”;
- **conversion** (or zero derivation), which is the process of creating a new word belonging to a different word class without adding any morphemes, e.g. “water” (a noun) – “water” (a verb), “spy” (a noun) – “spy” (a verb);
- **inflection**, which modifies a word to express different grammatical categories such as tense, mood, voice, aspect, person, number, gender, and case without changing the word class, e.g. the noun “sit” is a root for inflection forms “sits” and “sitting”;
- **modification** of a stem - can have a form of ablaut (a change of a vowel within the stem, e.g. *man* – *men*, *ring* – *rang*, *sing* – *song*); a consonant alternation at the end or the beginning of a stem (e.g. *believe* – *belief*); a change of stress: English *‘import* (noun) – *im’port* (verb); or a change of tone in tonal languages;
- **compounding** - combines two or more roots to create a new word, for example, “rattle” + “snake” = “rattlesnake” or “key” + “word” = “keyword”. The process is very productive in English, and other examples include words such as bookcase, doorknob, fingerprint, sunburn, textbook, wallpaper, wastebasket, waterbed, good-looking, low-paid, etc.
- **circumfixation** – is the process whereby a word is derived by a single morpheme split into prefixes and suffixes while neither of these affixes is used on its own. While frequently occurring in Germanic languages such as German and Dutch (see German past participles of regular verbs, e.g. *ge-spiel-t*), the process is rare in English, and literature usually mentions four examples: enlighten, embolden, enliven, and embiggen (Klégr, 2018). Circumfixation is sometimes wrongly confused with parasynthesis, a simultaneous attachment of a prefix and suffix as separate morphs to a root (prefix + root + suffix, e.g. un-comfort-able).

- **reduplication** – reduplication is a morphological process in which a root morpheme or part of it is repeated, e.g. in Indonesian, the word “orang” means a person and its plural form is “orang-orang” (persons); the Japanese word “yama” (meaning a mountain) after reduplication “yama-yama” marks mountains; a well-known example from Europe is a French word “bon-bon”, naming a candy, or literally “a good goodness” (for more examples see *Graz Database on Reduplication*, online; or Rubino, 2013).

Grammatical categories

There are many different grammatical categories of which individual languages make their choice. It means that the number and structure of grammar categories differ from language to language. Modern English, for instance, has over twenty grammatical categories, *including gender*, number, case, definiteness, person, tense, aspect, mood, voice, and degree.

Gender

Some languages distinguish gender as a grammatical category of nouns. The defining characteristic of gender is a syntactical agreement with other words (primarily verbs or adjectives). In Indo-European languages, gender is most often represented as a distinction between masculine and feminine or between masculine, feminine and neuter. Corbett (2013a) analysed 256 languages, of which over half (144) have no gender system. A minimal gender system consisting of two genders was identified in 50 sample languages. Three genders were recognised in 26 languages, and four genres in only 12 languages. Twenty-four languages distinguished five or more genders (cf. Table 5), including Mountain Arapesh with 13 genders (Roscoe, 2003), northern-Australian Ngan’gityemerri with 15 genders (Reid, 1997) and “the record-keeping” language of Nigerian Fula, having around twenty genders (Breedveld, 1995).






	Value	Representation
	None	145
	Two	50
	Three	26
	Four	12
	Five or more	24
Total:		257

Table 5. The overview of gender systems (source: Corbett, 2013a)

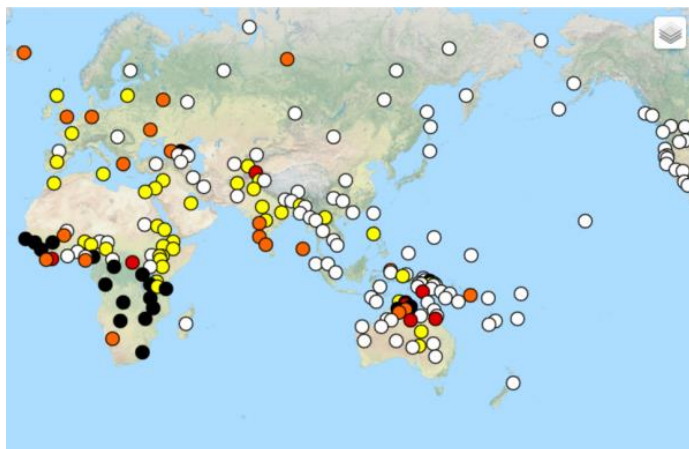


Figure 9. Number of genders in the languages of the world (Corbett, 2013b)

Linguistic gender systems are frequently linked to biological sex. In such a case, male humans are almost always assigned to the masculine gender, whereas women belong to the feminine gender. The grammatical phenomenon of sex-based genders was studied by Corbett (2013b) in *WALS*, Chapter 31. In his research sample, out of 112 languages with gender systems (see Table 5 above), 84 had sex-based gender systems, and the rest 28 had non-sex-based gender systems.

Number

In linguistics, *number* refers to the difference between singular and plural forms of nouns, pronouns, determiners, and verbs. “Number is the category through which languages express information about the individuality, numerosity, and part structure of what we speak about” (Acquaviva, 2017). However, not all languages mark this difference grammatically. Chinese and Japanese famously do not use plural forms of nouns. A complete lack of plural marking was documented in the aboriginal languages of Australia and New Guinea. Dryer (2013a) analysed the sample of 1066 languages for various ways of morphological indication of noun plurality and identified 98 languages with no plural (see Table 6).

On the other hand, many languages recognise more than two values in the grammatical category of number. Some languages, such as Arabic, have two plural forms, one for only two items and one for more than two. Others, along with singular and plural, also distinguish forms known as paucals (for small cohesive groups of something) or greater plurals (for large groups).

Value	Representation
● Plural prefix	126
● Plural suffix	513
● Plural stem change	6
● Plural tone	4
● Plural by complete reduplication of stem	8
● Morphological plural with no method primary	60
● Plural word	170
◆ Plural clitic	81
○ No plural	98
Total:	1066

Table 6:
Indicating
grammatical
number in 1,066
languages
(source: Dryer,
2013a)

Dryer (2013a) identified eight main ways of indicating noun plural (see Table 6) and apparent borders in their geographical distribution (Figure 10). The map shows that plural suffixes are the dominant way of expressing a plurality of nouns. Plural prefixes are frequently observed in Africa (especially Bantu and Niger-Congo languages). Other types of coding plural number, including plural by complete reduplication and specific plural words, are rare. Languages with multiple primary morphological methods for forming plurals are prevalent in North Africa, as well as spoken Arabic and Berber languages. Plural words and clitics are present in southeast Asia and among Austronesian languages. Finally, all four languages that use tone to indicate plural nouns are spoken in Africa.

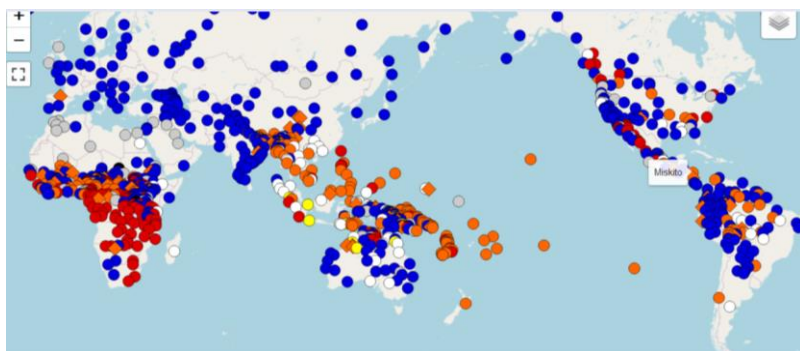


Figure 10. Geographical distribution of languages according to the ways they indicate grammatical number (source: Dryer, 2013a)

Case

A grammatical case expresses relations between nouns, noun phrases, or within nominal groups (consisting of a noun and its modifiers). “The semantic roles and grammatical relations indicated by the case are fundamental to the whole system of language and have long been a central concern of descriptive and theoretical linguistics” (Malchukov & Spencer, 2012).

Historically, the Indo-European languages had eight morphological cases (nominative, genitive, dative, accusative, ablative, vocative, locative, and instrumental). In the development of some languages, cases merged, which resulted in case syncretism (for more, see Baerman & Brown. 2013). Thus, in their modern forms, a reduced number of cases typically remained (e.g. six cases in modern Slovak and Czech, four cases in modern German and English, etc.). Grammatical functions of original cases have been substituted by articles, prepositions, adjectives, noun phrases, inflecting, and word order. In English, the initially complex and extensive case system has been reduced to personal pronouns only which still have three cases: subjective case (I, you, he, she, it, we, they, who, whoever), objective case (me, you, him, her, it, us, them, whom, whomever) and possessive case (my, mine; your, yours; his; her, hers; its; our, ours; their, theirs; whose; whomever).

Among the languages that keep extensive case systems are Greenlandic, Mongolian, Sanskrit, and Tamil, with eight cases; Basque, with thirteen cases; and Estonian, with fourteen cases. European languages with the most cases are Finnish, with fifteen cases and Hungarian, with 18 cases.

Definiteness

Definiteness is a grammatical category of noun phrases that refers to whether a noun is **a specific (identifiable)** entity. Definiteness can be coded in several ways. Dryer (2013b), who studied this phenomenon on a sample of 620 languages, identified those ways and frequency of their presence in languages as follows:

Value	Representation
Definite word distinct from demonstrative, e.g. English, German, Frech	216
Demonstrative words used as markers of definiteness, e.g. Latvian	69
Definite affix on noun, e.g. in Arabic, Danish and Swedish	92
No definite article but they have indefinite articles, e.g. Turkish and Japanese	45
Neither definite nor indefinite article, e.g. Czech and Slovak	198
Total:	620

Table 7. Expressing definiteness (source: Dryer, 2013b)

English is an example of the first category since it codes definiteness utilizing the definite article “*the*”, distinct from the demonstratives *this/these* and *that/those*.

Person

The category of **person** is concerned with participation in the speech act. In inflectional languages, the system of persons is often combined with the grammatical category of number (singular vs. plural). It is formally marked in personal pronouns (*I, you, we* ...) and/or in inflectional forms of verbs. In this perspective, most languages distinguish three categories: **first person** = the speaker (*I/we*), **second person** = the addressee (*you*), and **third person** = other persons or things not participating in the speech act (*he, she, it, they*). Some languages are strict in expressing whether “*we*” means “*I and you*” or “*I and somebody else but not you*”. This interesting distinction, absent in European languages, was studied by Cysouw (2013), who mapped the occurrence of this inclusive versus exclusive distinction in independent pronouns. On the sample of 200 languages, he observed that 63 languages recognised such an opposition (including Mandarin Chinese or Comanche) while most languages (120) did not. Ten languages use the same pronoun for expressing „*I*“ and „*we*“ concepts (e.g. Vietnamese). Five languages had a special pronoun for the inclusive „*we*“, but the expression of the exclusive pronoun was identical to “*I*“ (e.g. Mexican Mixtec).

Value	Representation
No grammaticalised marking at all	2
'We' and 'I' identical	10
No inclusive/exclusive opposition	120
Only inclusive differentiated	5
Inclusive and exclusive differentiated	63
Total:	200

Table 8. The occurrence of this inclusive versus exclusive distinction in independent pronouns (source: Cysouw, 2013)

Tense

A grammatical category of tense expresses when an action took place and is linked to verbs. The main verb tenses are the past, present, and future, systematically distinguished by various grammatical means in most Indo-European languages. Yet, as pointed out by Dahl and Velupillai (2013a), many languages lack any grammatical distinction between past and present, e.g. the Indonesian sentence *Air itu dingin* means both „The water is cold“ and „the water was cold“.

Some languages do not have grammaticalised future tense, i.e. expressed by a specific grammatical change of a verb such as inflection (in Slovak *utekám* – *utečiem*). English is one of them. However, it is relatively rare for a language to entirely lack

grammatical means for marking the past or the future. Most languages have at least one or more grammaticalised devices for doing so, e.g. through auxiliaries and adverbs (Dahl and Velupillai, 2013b).

For native speakers of Indo-European languages, it might be fascinating that many languages use several past tenses depending on the temporal distance between the time of speech and the time of action (*remoteness distinctions*). Within the research carried out by Dahl and Velupillai (2013a) on the sample of 222 languages, the wealthiest system of past tenses had Yagua (the language used in Colombia and Peru), which had five degrees of remoteness (see Table 1, c.f. also Payne & Payne, 1990).

Name in grammar	Use	Suffix	Example
Proximate 1	a few hours previous to the time of utterance'	-jásiy	<i>rayáásiy</i> {ray-jiya-jásiy} 1sg-go-prox1 'I went (this morning).'
Proximate 2	'one day previous to the time of utterance'	-jay	<i>rjijnúújeñíí</i> {ray-junnúúy-jay-níí} 1sg-see-prox2-3sg 'I saw him (yesterday).'
Past 1	'roughly one week ago to one month ago'	-siy	<i>sadííchimyyaa</i> {sa-díí-siy-maa} 3sg-die-pst2-perf 'He has died (between a week and a month ago).'
Past 2	'roughly one to two months ago up to one or two years ago'	-tíy	<i>sadíítimyyaa</i> {sa-dííy-tíy-maa} 3sg-die-pst2-perf 'He has died (between 1 to 2 months and a year ago).'
Past 3	'distant or legendary past'	-jada	<i>raryúpeeda</i> {ray-rupay-jada} 1sg-be.born-pst3 'I was born (a number of years ago).'

Table 9. Remoteness distinctions in Yagua (source: Dahl and Velupillai, 2013a)

Aspect

Aspect as a grammatical category expresses how an action, event, or state represented by a verb extends over time. **The perfective aspect** indicates a one-time action, while the **imperfective** aspect points to a continuous or habitual action.

In their research, Dahl and Velupillai (2013c) observed that “the distinction between **imperfective** and **perfective** plays a vital role in many verb systems and is commonly signalled by morphological means“. They divided languages from around the globe (N = 222 languages) into those with the grammatical marking of the perfective/imperfective distinction (either by morphological means or by periphrastic constructions) and those where there is not. The results are represented in Table 10).

Value	Representation
Grammatical marking of perfective/imperfective distinction	101
No grammatical marking of perfective/imperfective distinction	121
Total:	222

Table 10. Languages with and without grammatical marking of perfective/imperfective distinction (source: Dahl and Velupillai, 2013c)

Grammatical mood

In linguistics, mood is a grammatical feature of verbs which indicates the speakers' attitude towards what they are communicating. The speaker may wish to state a fact (an **indicative mood**), learn a new piece of information (**an interrogative mood**), or order somebody to act (**an imperative mood**). Other types of mood are conditional, subjunctive, injunctive, optative, and potential. In English, only verbs in finite forms can express the grammatical mood, while non-finite forms of verbs, such as infinitives, gerunds, and participles, cannot.

Voice

The grammatical voice (diathesis) marks the relationship between the action (or state) and its participants. The verb is in the active voice when the subject acts. When the subject is the receiver, target or undergoer of the action, the verb is in the **passive voice**. Some languages recognise the **middle voice** when the subject is a performer and a target of the action (Zúñiga & Kittilä, 2019).

Degree

A grammatical degree is the category of adjectives and adverbs which indicates their quality or intensity. Languages usually distinguish three degrees of comparison. The basic form of an adjective or adverb, the **positive degree**, shows the quality of the noun it determines. A **comparative degree** expresses a higher level of intensity. A **superlative**

degree marks the greatest quality or intensity compared to other comparators (see Hohaus & Bochnak, 2020).

Word classes

Words with the same grammatical characteristics belong to the same **word class**. Not all languages have the same word classes, but many share common categories. The most common word classes include nouns, verbs, adjectives, adverbs, pronouns, prepositions, and conjunctions. Some languages may also have additional classes, such as classifiers (Mandarin Chinese), particles (Slovak), and postpositions (Turkish and Uralic languages).

Grammar textbooks of modern English usually recognise between eight and eleven word classes. The maximalist list of English word classes includes nouns, pronouns, verbs, adjectives, adverbs, prepositions, determiners, conjunctions, interjections, articles and numerals. Many English words belong to more than one-word class, and their actual categorization depends on their position and function in the sentence, e.g. *fast* can be used as an adjective (*a fast runner*) or an adverb (*run fast*). However, in many other inflectional languages, a given word (in its concrete form) can be identified as part of only one specific word class with related grammatical properties (like in Slovak).

Syntax

Syntax studies arranging words into phrases, clauses, and sentences. Hypersyntax studies the rules of organising sentences into paragraphs and longer texts.

In most languages, the simplest form of a sentence consists of a noun phrase (which might be a noun) + a verb phrase (which may be a single verb).

Essential parts of the sentence

In most languages, every word in a sentence usually serves a specific syntactic purpose. According to this purpose, words are divided into several parts of the sentence. In various languages, the number and structure of parts of a sentence vary. In English, the following parts of speech are determined: subject, predicate, (direct) object, indirect object, and subject complement are named.

The **subject** is the person, place, or thing performing the sentence's action. The simple subject most frequently contains a noun or pronoun and can include modifying words, phrases, or clauses, creating a noun or pronoun phrase.

The teacher explains grammar to students.

The **predicate** expresses action. The simple predicate contains the verb. The predicate phrase also contains modifying words, phrases, or clauses.

The teacher explains grammar to students.

In English, a full sentence can only be composed of a simple predicate, usually in imperative sentences. Thus, the shortest full English sentence is “Go!”

The **direct object** (usually in the form of a noun or pronoun) receives the action of the sentence.

The teacher explains grammar to students.

The **indirect object** indicates to whom, for whom, or with whom the action of the sentence is being done. The indirect object is usually a noun or pronoun preceded by a preposition. Only transitive verbs can be linked to both direct and indirect objects.

The teacher explains grammar to students.

A **subject complement** describes or adds details to the subject. It is, similarly to a subject, usually a noun, pronoun, or adjective. The unmistakable signal of the subject complement’s occurrence is a **linking verb** (be, become, seem, look, make, etc.)

The students make a great audience.

Her children were dreadful students.

The previous model sentences illustrate the difference between parts of speech and parts of a sentence in English. While the word “students” is always a noun in plural form, it functions within two sentences differently - as a subject and a subject complement. The same noun can also efficiently function as a direct or indirect object.

Sentence structure

In syntax, a group of words containing a subject and a verb is called a clause. If a clause makes a complete sentence, it is an **independent clause**.

She does not appear very happy.

Another type of clauses which cannot stand independently (therefore, they are named **dependent clauses**) but they add necessary information to or support independent clauses are marked as the subordinate clauses.

She does not appear very happy because her shoulders are slumped, her eyes seem distant, and she hasn't smiled all day.

This sentence contains four clauses:

- (1) *She does not appear very happy,*
- (2) *because her shoulders are slumped,*
- (3) *her eyes seem distant,*
- (4) *and she hasn't smiled all day.*

Each sentence has its subject and a verb, but only the first clause is a complete sentence, while the remaining three sentences add explanation details.

Based on various arrangements of clauses within a sentence, four different types of sentence structure can be distinguished:

- **simple**, consisting of one independent clause
- **compound**, consisting of two or more independent clauses linked by a compounding conjunction
- **complex**, composed of one independent clause and one or more subordinate clauses
- **and compound-complex sentences**, which consist of two or more independent clauses and one or more subordinate clauses.

Word order

The most obvious signal of syntax is word order. Contemporary linguistics recognizes six logical possibilities for the word order of a simple sentence: SOV, SVO, OSV, OVS, VSO, and VOS (where S = subject, V = verb in its finite form, O = object). Dryer (2013c) analyses the frequency of these word orders within the sample of 1376 languages. He discovered that the most frequent of the six structures is SOV, followed by SVO. In addition, they both are widely distributed across the globe. Word orders OVS and OSV were identified as much rarer (see Table 11).

Value	Representation
Subject-object-verb (SOV), e.g. Turkish, Japanese, Korean	564
Subject-verb-object (SVO), e.g. English, Chinese, French	488
Verb-subject-object (VSO), e.g. Irish, Welsh, Arabic	95
Verb-object-subject (VOS), e.g. Malagasy (in Madagascar), Dusun, Mayan languages	25
Object-verb-subject (OVS), e.g. Guarijio, Urarina	11
Object-subject-verb (OSV)	4
Lacking a dominant word order	189
Total:	1376

Table 11. Order of subject, object and verb in the world languages (source: Dryer (2013c))

Some languages prefer only one type of word order, considered the only grammatical one (e.g., English). They assign other word orders as ungrammatical, non-standard, or stylistically marked, i.e. appropriate only in unique communicational contexts. Such languages are known as having **rigid word order**. Other languages (including Slovak and Czech) approbate all six types or at least several types of word order as grammatical. They have a **flexible (or free) word order**. However, the choice of actual word order is never random or chaotic but governed by communication purposes and pragmatic factors.

7 Meaning in language

Study objective(s): After discussing the grammatical chapter, this chapter debates the lexical meaning of words. The key terminology of lexicology, i.e., the linguistic discipline that studies verbal lexical meaning, is introduced in the first part. Next, the concepts of lexicons and the size of vocabularies of natural languages are outlined. Then, semantic shifts, changes in words and lexical relations among words are considered. The last part of the chapter focuses on dictionaries and their creation, which is the process analysed by lexicography.

Linguistic disciplines: general linguistics, lexicology, lexicography, anthropological linguistics, psycholinguistics

Key terms: word, lexical meaning, lexeme, arbitrariness, denotation, connotation, collocation, lexicon, basic lexicon, metaphor, metonymy, corpus, dictionaries, digital dictionaries

7.1 Words and lexemes

The primary purpose of language is to share meaning and to be meaningful. The study of verbal meaning is probably the most challenging part of linguistics due to the nearly never-ending number of variables that construct the final meaning of a verbal utterance - communicated message. The smallest unit that can transfer a meaning (a message) is called **a word**.

The meaning of the word consists of two components: grammatical and lexical. Lexical and grammatical meanings of the word are coexistent and usually inseparable. While the grammatical component is dominant in some words, lexical meaning prevails in other (much larger) groups of words. However, “even when lexical words bear no morphological markers of the grammatical meaning, the grammatical component of meaning is still present” (Pavlík, 2017, p. 73).

The grammatical meaning of the word is expressed by a specific pronunciation pattern, a particular manner of word formation, its grammatical categories, its ability to be combined with other words, and its placeability in the sentence. Even though phonemes and morphemes, essential components of grammatical meaning discussed in previous chapters, may bear and change verbal meaning, they cannot convey a whole message.

The lexical meaning of the word reflects the content of communicated information. Within linguistics, the lexical meaning is studied by **lexicology**. It deals with the study of properties, usage and origin of words. It includes the study of naming extra lingual reality (*onomasiology*), the study of verbal meaning (*semasiology*), the study of the history of

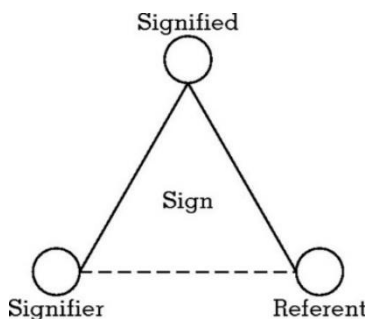
words (*etymology*), the study of word formation (*lexical morphology*), and the study of lexical phrases, idioms (*phraseology*). The vital role within lexicology is played by **lexicography**, the theory of organizing lexemes in dictionaries. Because of shared research in processes that create and understand the meaning, lexicology closely cooperates with psycholinguistics. The study of word meaning is crucial to understanding language and the knowledge of the nature of human thinking and being. Therefore, the word's meaning is studied by linguistics, developmental psychology, and philosophy, especially by its specific branch, **semiotics**.

Lexicology is the linguistic discipline that studies a specific language's lexicon (vocabulary). A lexicon is the sum of all words which belong to or are used in an individual language. Every lexicon consists of **lexemes**, which are “the smallest linguistic units that can be articulated in isolation to convey semantic content. Just as phonemes are abstract representations of sets of phones (...), lexemes can be defined as abstract representations of sets of words (each defining one way the lexeme can be instantiated in the appropriate sentence environments)”. Thus, concrete words used in speech, 'do', 'does', 'did' and 'doing' are in human speech concrete realizations of the abstract lexeme 'do' (Gasparri & Marconi, 2024). Lexemes are the result of mental analysis and categorization of the extra-lingual reality. The processes of creating, storing, structuring, and using items from the mental lexicon (personal vocabulary of a speaker) are studied by **psycholinguistics**.

The arbitrariness of lexical meaning

In the second chapter, arbitrariness was mentioned as one of the leading human language properties. It marks the indirect relationship between the word's form and its meaning. Many experts have tried to explain this relationship in numerous theories. Ferdinand de Saussure proposed one of the most respected theories in the 19th century. He indicated that when humans communicate by words, three things come into play: a physical object (*the referent*, e. g. a table), the abstract concept in human minds which refers to that piece of furniture (*the signified*), and a word which symbolises that concept in a particular language (*the signifier*, "a table"). This process (*semantisation*) enables people to speak about remote places and events or create newly constructed reality (stories or lies). Saussure modelled his explanation by a semantic triangle (see Figure 11).

Figure 11. The Saussure's semantic triangle



The principle of arbitrariness, i.e. the lack of a direct relationship between forms and lexical meanings, is characteristic of most words in any natural language (note: a grammatical meaning of verbal structures is arbitrary, too).

A few exceptions are onomatopoeic words, the sound forms which express their meaning (sound symbolism). Due to this direct relation, some onomatopoeic words are similar in various languages, e.g. English *cuckoo*, Slovak *kukučka*, Czech *kukačka*, French *coucou*, and German *Kuckuck*. However, other onomatopoeic words might be very different, e.g. compare the sounds produced by pigs as heard in various languages:

English: oink

French: groin-groin

Hungarian: rőf-rőf

Polish: chrum chrum

Slovak: kroch kroch

Spanish: grrr

Japanese: ブー(buu)

Hindi: ओई ओई (oi oi)

Arabic: خنخنة khankhanah

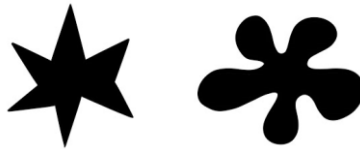
Chinese: 哼哼 (hēng hēng)

(source: Zakryzhevskyy & Mendes, 2024)

In addition, some neurolinguists are working on the hypothesis that human language at least partially originates in the human ability of synesthesia, i.e. an intuitive process which involves the crossing of different senses. The idea was tested by a so-called “Bouba/Kiki effect” experiment (Ćwiek et al., 2021). Within it, researchers showed nearly a thousand people of 25 languages representing nine language families and ten writing systems in two different shapes (see Figure 12). They asked them to decide which of them is called Bouba and which one is Kiki. Regardless of their language or cultural background, 96% of respondents believed that Bouba is related to the rounded shape, while Kiki is linked to the pointed shape. In this way, these links are recognised by nearly all people as violating the arbitrariness principle. The researchers argued that the

“Bouba/Kiki effect” derives from phonetic and articulatory features of the words, i. e. the rounded lips of the /b/ and the stressed vowel in “bouba”, and the intermittent stopping and starting of air in repeated pronouncing of a plosive /k/ in “kiki”.

Figure 12. Bouba and Kiki



7.2 The size of the lexicon

Every language's vocabulary consists of smaller or larger sets of words (from hundreds to millions). From time to time, popular literature publishes charts of the most robust languages (i.e. languages with most words in their lexicon). Some sources put English at the top (Andrews, 2024), others Korean (Wood, 2022), or modern Arabic (Lonet Academy, 2024). Such charts may be attractive to readers; however, their validity is questionable. Since lexicons are ever-changing and precarious systems, it is nearly impossible to state the number of words in any national lexicon.

Lexicons grow with every new word their users create and use effectively, which happens every day. They may not be as prolific as William Shakespeare, who is said to have invented at least 1700 words, making English a successful literary language.

Most words in Indo-European languages have been created and continue to be created by lexical word-formation processes that have been explained in the previous chapter as morphological processes: **derivation** (including affixation, prefixation, suffixation), **conversion**, **modification**, **circumfixation**, **compounding**, and **reduplication** (note that inflection does not create new lexemes; therefore, it is not dealt with in lexicology, c.f. Pavlík, 2017). Other productive word-formation processes include:

- **clipping**, in which a new word (often informal) is formed by shortening a longer word, e.g. ad (advertisement), exam (examination), math (Mathematics), chem (chemistry), lab (laboratory), gym (gymnastics or gymnasium), etc.
- **blending**, in which parts of two or more words are combined to form a new word, e.g. „brunch“ (breakfast + lunch) or „smog“ (smoke + fog“).
- **initialisation**, the products of which are acronyms or initialisms formed from the initial letters of a phrase. While initialisms are pronounced letter by letter (e.g. USA, CIA, OMG, LOL), acronyms are pronounced as words (laser = light amplification by the stimulated emission of radiation, GIF = graphics interchange format, NATO = North Atlantic Treaty Organisation).

Speakers can borrow words from other languages, too, enriching thus the repertoire of their language (see **loanwords** and **borrowings**, e.g. piano, bona fide, tête-à-tête).

In time, some words lose their appeal and may eventually be suppressed or forgotten. However, lexicology does not recognise or describe the process of deleting or excluding a word from a particular vocabulary). Instead, it recognises temporary obsolete words such as **archaisms** (words or expressions that are not generally used anymore; they can be replaced by more modern tools, e.g. *thy* - *you*) and **historicismisms** (words related to objects, events, processes, and concepts that do not exist anymore).

Instead of identifying the language with most words, linguists find it more productive to search for the **minimal lexicon**, the smallest set of lexemes sufficient for effective communication. For these experts, languages with limited vocabularies are inspirational, either native languages such as Nahuatl (a language in Mexico of something over 10,000 words) or artificially constructed languages such as Toki Pona with 120-260 words (see Toki Pona, online).

For decades, linguists involved in comparative linguistics and lexicostatistics have tried to determine the **basic lexicon** - a list of forms and concepts for which all languages have terms (words). Among the expected items are names of body parts (*head, eye, blood*), everyday human activities (*eat, drink, see, walk*), natural phenomena (*sun, water, fire*), small numbers (*one, two, three*), and personal pronouns (*I, you, he*). One of them was the American linguist Morris Swadesh, who published his first list of universals nearly 75 years ago (Swadesh, 1950), inspiring many followers today. The methodological weakness hidden in such lists is that they are created primarily as a collection of words in a particular large language (most often English) and then complemented by translations to other languages. A much more appropriate method is to define topics (e.g. colour terms, kinship terms, temporal terms, etc.) and then collect and organise the concrete means of expression in individual languages. These days, the method is used for the ASJP database (Wichmann, Holman, & Brown, 2022) or IDS project (Key & Comrie, 2023).

7.3 Semantic shifts and changes in words

Another complication related to the determination of the exact size of the vocabulary of any natural language is the fact that the lexical meaning of words is prone to various changes or semantic shifts. Over time, a particular word either loses some of its components of meaning or acquires new components. Contemporary lexicology recognizes two main types of such changes: specialization (or narrowing) and generalization (broadening).

The following principles represent these changes:

- **metaphor** - the new meaning of the word is drawn from the older one based on similarity or analogy;
- **metonymy** - the word's new meaning is derived from the older one based on some inner quality. Metonymy is the umbrella term for subordinate principles:
 - (a) **synecdoche**: a part of something is used to name the whole or vice versa, e.g. a city (= inhabitants of the town)

(b) **synecdoche of origin**: something is named after the material it is made of, e.g. an iron (= a tool for pressing clothes);

(c) **containment**: when one thing contains another, e.g. a dish (= the food in a dish).

(d) **eponymy**: a new meaning of the common word is related to the proper names of people or countries, e.g. sandwich, boycott, china, etc.

- **personification**: new lexeme originates in assigning human qualities to non-human entities, e.g. a face of the clock;
- **euphemism** and **dysphemism** refer to substituting a neutral expression with a mild, conforming (euphemism, e.g. using the phrase “a senior person” instead of “an old person”) or negative, disturbing expression (dysphemism, e.g. naming a person a snake).

Note that the branch of linguistics which studies the origin of words and changes in their form and meaning they took over time is called **etymology**.

7.4 Lexicon as a reflection of the needs of the community

A linguistic repertoire of an entire language enables its users to express everything they need. Understandably, different speech communities have different communicative needs, reflected in differences in their vocabularies. For example, vocabulary related to computers and digital technology is present only in the languages of those communities that use it. Otherwise, incorporating the words naming parts of the computer or verbs marking various ways of virtual communication would be useless. On the contrary, in most European languages, transporting an object while holding it on some part of one's body is usually named by one verb – *carry* in English, *niesť* in Slovak, etc. The importance of the activity for other communities is accented in their languages. For instance, speakers of ǀAkhoe Hai||om language (used in Namibia) could choose from a much wider repertoire of verbs naming the same action (cit. from Nau, 2014):

Ton	‘carry on one’s shoulder’
!guri	‘carry on one’s head’
gobe	‘carry on one’s back’
aba	‘carry (a baby) on one’s back.’
ǀkhore	‘carry a load’

Popular and expert literature offers similar examples that manifest the **cultural determination of the lexicon**. In Hawaii, speakers have no general word for weather, although they can choose from around 150 words that mark various types of rain. Some African languages have separate words for the left hand and the right hand but no word for the hand in general (Pavlík, 2018). In Pirahã language (one of the tribal languages in

Brazil), the absence of words for numbers and colours keeps surprising linguists (Everett, 2005). While many European languages have developed an extensive vocabulary for cattle, in Baka (the language used in Central Africa), many different words refer to elephants, usually represented by just one word in European languages (Paulin, 2010; cit. in Nau, 2014):

Baka word	Explanation
ijà	elephant in general
ndzàbò	giant male elephant, king of elephants
sèmē	old big male elephant
kàmbà	big male elephant (but not as strong as the above)
mòsèmbi	male elephant (between kàmbà and mòbòngɔ)
mòbòngɔ`	smaller male elephant
èkwāmbē	young male elephant living alone; male or female elephant that has lost its mother and become solitary
lìkòmbà	female adult elephant
bèndùm	elephant calf

Cultural determination of vocabulary is evident in the area of kinship terms. These words not only name family members but also reflect on the social structures within communities and the relationships between their members. This is why the structure of kinship terms is closely studied in **cultural anthropology** and **anthropological linguistics**. English distinguishes the nearer kinsfolk by sex: *mother, father, sister, brother, aunt, uncle*, etc. Other languages make lexical distinctions based on age, using separate words for elder brother or sister and younger brother or sister. Some languages differentiate terms for father's and mother's brother and use different terms for their children. In contrast, in other languages, both mother's and father's brothers are marked by the exact words (e.g. *uncle* in English), and they do not use different words for their children (in English, they all are *cousins*).

In Hawaiian languages, male cousins are referred to by the same term as brothers (*kaikua'ana*) and both female cousins and sisters are marked by one word (*kaikuahine*). Hawaiians do not recognise the difference between mother and aunt or father and uncle.

Psycholinguistic aspects of creating word meaning are visible when studying number terminology, i.e. words used to mark distinct quantities. Native speakers of many European languages would believe that words naming numbers up to 20 belong to the core vocabulary of all languages. However, based on the different communicative needs of communities and the prevalent ways they perceive their surroundings, their strategies of creating and using number words may be surprisingly diverging.

Both cultural and anthropological linguistics proved that all human languages have the means to express the idea of quantity. Most languages use a decimal counting system with a base of 10 (= the number of fingers). Similarly, a large group of languages use a vigesimal system based on 20 (= number of fingers and toes). Other languages use other number bases, such as base four (Tonkawa, many Papua New Guinean languages), base eight (Yuki and Pame), base twelve (many Nigerian languages), or base sixty (Sumerian). Very special in this regard are languages with body-part counting systems which express the exact quantity of something by using the name of the related body part. Words for numbers are thus identical to words for specific body parts (e.g. in Oksapmin, one of Papua New Guinea's languages).

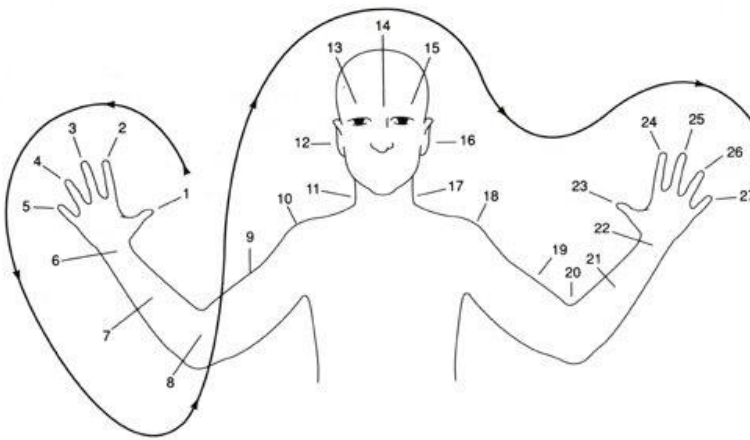


Figure 13: Oksapmin body-part counting system (source: Saxe & Esmonde, 2004)

Other human languages do not have specific words for numbers since their speakers do not engage in counting (e.g. hunter-gatherer societies). They do not need numerals and can express quantity in other ways. For instance, the language Pirahã has only three words that might be called numerals: *hói* (small number/one), *hoi* (a larger amount/two), and *báagiso* (many/three).

7.5 Relations among words

Contemporary lexicology claims that the vocabulary of any natural language is not a mere collection of individual words and phrases. On the contrary, it is a system comprising complex relationships between lexemes and networks. Words may be related in terms of their semantic, stylistic, social, temporal, geographical, and other characteristics.

- **Semantic relations** are given by mutual links in the meaning of words. The essential semantic relationships are synonymy (the similarity of meanings), antonymy (the oppositeness of meaning), polysemy (the multiplicity of meaning), and semantic fields (sets of words with the same denominator of meaning).
- **Temporal relations** tie together lexemes that originated in a particular period. They explain the differences between archaisms, contemporary vocabulary and neologisms.
- **Geographical relations** mark lexemes related to a particular geographical area and put words from different locations into opposition, such as the words differently naming the same entities in British and American English, e.g. lift and elevator.
- **Sociolinguistic relations** are given by the fact that members of different social classes or groups usually use specific vocabulary, e.g. teenage slang and medical jargon.
- **Stylistic relations** connect the words belonging to particular stylistic levels. For instance, formal, neutral and informal words are stylistically interrelated (Pavlík, 2018).

7.5 Lexicography

Lexicons of languages are studied by **lexicography**, the theory and practice of compiling dictionaries. Lexicography is concerned with the systematic selection, compilation, description, and organization of words in dictionaries. Word descriptions and characterizations in general dictionaries (i.e. dictionary entries) usually include information on pronunciation, meanings, grammatical structure, stylistic characteristics, word usage, etc. Dictionaries list these words and describe their meanings (denotative, connotative, grammatical and others), applying selected linguistic criteria. Each word in the utterance can carry several types of meanings (c.f. Leech, 1981), including:

Denotative meaning (also fundamental, conceptual or referential) is the word's primary, core meaning. Other meanings refer to it.

Connotative meaning (also associative or supplementary) refers to the additional, non-criterial properties of the word, frequently based on association or various types of semantic shifts (see subchapter 7.3). Connotations do not affect the lexeme's essential components of meaning.

Grammatical meaning includes morphological and syntactic information about the word. Morphological information refers to the grammatical category of a word (noun, verb, adjective, etc.), the gender and/or declension class of nouns, the conjugation class of verbs, and any irregularities of form when a word is inflected for number, gender, and person. Syntactic information includes the contexts in which a word typically occurs, e.g. an adjective usually precedes a noun).

Expressive meaning (affective, emotive, attitudinal) points to the personal feelings and attitudes of the speaker.

Social meaning expresses how language reflects societal norms, attitudes, and hierarchies, e.g., in educated speech, jargon, slang, etc.

Thematic meaning is related to the central theme or topic of the sentence.

The collocational meaning of the word encompasses its ability to be combined with other words and to form collocations (relatively literal word combinations) or idioms (metaphorical phrases). It reflects how the meaning of the lexeme is changed in the neighbourhood of other words. The linguistic discipline that investigates formal and semantic aspects of word combinations is called **phraseology**.

According to the scope of their wordlist, linguistic dictionaries can be divided into general and specialised. General dictionaries aim to represent the vocabulary with a defined degree of completeness (e.g. *The Oxford English Dictionary*, *Krátký slovník slovenského jazyka*), etc. General dictionaries usually function as manuals of standardised language.

Specialised dictionaries describe only a particular specific part of the vocabulary. Well-known examples are medical dictionaries, dictionaries of technical terms, dictionaries of linguistic terms, etc.

According to their various functions and the types of meanings they cover or prioritize, the following types of dictionaries can be distinguished:

- *monolingual dictionaries* where the words and the information about them are given in the same language;
- *bilingual dictionaries*, primarily used for translating and foreign language learning, explain words by giving their equivalents in another language;
- *picture dictionaries* which combine verbal and visual explanations of words;
- *dictionaries of synonyms and/or antonyms*, *dictionaries of collocations*, *dictionaries of idioms*, etc. - are dictionaries intended primarily for language professionals;
- *historical* and *etymological dictionaries* – organise and explain words according to their origin and evolutionary changes over time based on a systematic study of changes affecting a lexical unit during its life.

The scope of dictionary types is unlimited. They also include dictionaries of abbreviations, spelling or orthographical dictionaries, pronunciation dictionaries, word formation dictionaries (including dictionaries of roots, verbs, etc.), grammatical

dictionaries, reverse dictionaries, dictionaries of homonyms, dictionaries of paronyms and acronyms, dictionaries of idioms, etc.

Digital dictionaries

Digitalization has significantly impacted lexicography on several levels. It helped to create and work with much larger lexical databases with rich, multi-faceted, and ever-flexible representations of word meaning. Access to large-scale **electronic text corpora** (i.e. collections of vast numbers of texts) provides an attractive base for linguistic research since they allow the lexicographers to design much more effective dictionary entries which reflect actual speaker usage, variations across genres and the dynamics of continuous language change. Digitalization led to the development of new formats of dictionaries, i.e. **digital dictionaries**, which became crucial for the development of other digital tools, such as **machine translators**.

Compared to printed dictionaries, the undisputable advantages of digital dictionaries and machine translators are their flexibility (openness to be continuously updated) and their ability to connect, interlink and harmonise with other digital tools (c.f. Fellbaum, 2014).

8 Writing systems

Study objective(s): Humans invented various writing systems to share and store valuable information and keep records of important events. The chapter discusses their history, basic units, and classification based on their relationship to language. The last subchapter briefly outlines differences between spoken and written languages, which is the area studied by discourse analysis.

Linguistic disciplines: historical linguistics, discourse analysis

Key terms: writing system, grapheme, orthography, literacy, pictograms, pictographs, semasiographs, idiographs, logographs, syllabic writing, phonemic writing, alphabet, character script

As discussed and explained in Chapter 2, it is generally believed that language initially existed in the form of speech (oral communication). Everything that people recognised as important enough or needed to be remembered in the collective memory of a community was transmitted from generation to generation by oral literature. People sought more practical, stable, and lasting means of recording language, which led to writing development.

Writing is the symbolic representation of language through visual symbols called **graphemes**. Understandably, there is not only one way of writing spread worldwide; instead, humans developed multiple writing systems. A **writing system** is a way of recording a language established with a set of visual symbols and rules of **orthography**.

Similarly to the situation with a multiplicity of languages, many writing systems were used from the oldest times of human history (note: the times before writing are called pre-historic). While some of them died out and are not used anymore, others have been developing through these days.

The vast majority of languages were initially spoken, and much later, writing systems were used. However, a few languages had only a written form (so-called literary languages), such as Classical Tibetan, Classical Chinese wenyan, and kawi, a literary language used in Java (cf. Wicherkiewicz, 2014).

Writing is one of the communicative skills, and reading must be learned through intentional training. The ability to use a writing system (through reading and writing) is called **literary**. People who lack the skills are *illiterate* or *analphabets*.

History and classification of writing systems

Some experts recognise cave paintings (as graphic narratives) and various long-lasting human-made visible marks used to store information dated back as far as 20

thousand years as the first writing attempts. First, proto-writings (ancestors of written texts) are preserved as images cut to bones or shells, cave drawings, decorations on pottery items, knots or seals to mark property, etc. (Robinson, 2009). The images they used are **non-linguistic pictograms** because they did not refer to language units (phonemes, morphemes, syllables, or words) but to mental representations of objects or ideas. They are called **semasiographs** and are studied by semasiology. **Semasiographic systems** are often evaluated as not “full writing” because they strongly depend on an existing context, and their use and communicative effectivity is somewhat limited (Coulmas, 1996, 2003; Powell, 2012; Zadka, 2019).

Non-linguistic semasiography predates linguistic (language-based) writing systems, called **glottographic systems**. They are further divided into logographic and phonographic groups (see Figure 15). In **logographic systems**, graphic signs refer either to ideas (ideographic writing, such as ancient Egyptian hieroglyphs and contemporary Chinese character script) or to individual words (**pictographic writing**, used by the most ancient writing systems of Mesopotamia, the Inuits, and some contemporary artificial languages such as Toki Pona). The oldest generally attested form of writing, Sumerian **cuneiforms** carved on stone or clay tablets, also belong to pictographic writing systems. Sumerian cuneiforms were visual symbols that, similarly to semasiographs, encoded meaning by images visually resembling real-world objects, i.e. the picture of a bird referred to a bird, or an image of a mountain meant a mountain. The cuneiform script was used to write many extinct languages, e.g. Sumerian, Akkadian, Old Persian and Ugaritic. A key weakness of pictographic writing systems is that they typically consist of many symbols, and still, they can hardly express all the abstract ideas one needs to express. That is probably why people continued designing other writing systems. However, communication by pictograms is relatively frequent and widespread today, too; e.g. think about road signs, Olympic sports pictograms, no-smoking signs, emojis, etc. (see Figure 14).



Figure 14. Examples of modern pictograms

In **phonographic systems**, graphemes refer to sound units of language. In that case, graphemes refer to syllables (**syllabic writing**, such as in Japanese, Cherokee, and

Dravidian languages in South India) or phonemes, as in **alphabetic writings** (e.g. English and Slovak). Some languages developed a specific variant of an alphabetic writing system where only consonants are represented - **consonantal writing systems** (e.g. Arabic and Hebrew). Based on results by Comrie (2013), who researched various writing systems and their geographical distribution, the alphabetic writing system is dominant worldwide. Their strength lies in that they can, in stark contrast to ideographic writing systems, express a limitless amount of ideas using a limited repertoire of graphemes. Even the most extensive alphabet, used in the Cambodian language, Khmer, consists of 74 characters. The Hawaiian writing system consists of only 13 letters, which is still one letter more than Rotokas, which uses only 12 characters.

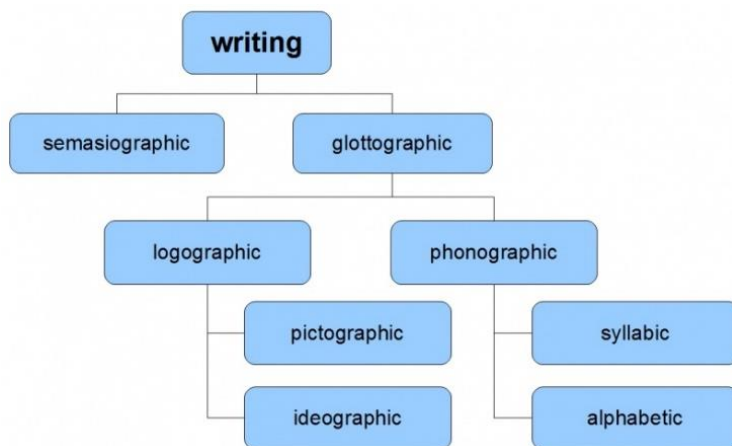


Figure 15.
Classification
of writing
systems
(source:
Wicherkiewicz,
2014)

Pure forms of writing are difficult to find since languages usually combine features of more types. For instance, ancient Egyptian hieroglyphs represented the objects they depicted and particular (typically initial) sounds or groups of sounds. The Egyptian hieroglyphic writing system consisted of more than 1,000 distinct characters that combined ideographic, logographic, syllabic, and alphabetic elements. Another example is contemporary Chinese character script, where individual characters stand for Chinese words or parts of words (*logograms*); some stand for abstract ideas (*ideograms*), but many still bear marks of pictorial origin (*pictograms*). Japanese is an example of a **mixed writing system** since it uses logographs to represent lexical morphemes (e.g. nouns, verb roots) and syllabary (i.e. syllabic graphemes) to represent grammatical morphemes, such as suffixes and particles (Comrie, 2013).

Written versus spoken language

Initially, linguists considered written language only a variant of spoken language. However, in recent decades, written language has been defined and studied as a separate system with specific norms, structures, and stylistic conventions. These differences between spoken and written language are examined by a linguistic discipline named **discourse analysis** (c.f. Ferenčík, 2016).

As Brown and Yule (1983) have it, spoken and written language make different demands on language producers. Generally, spoken language, which dominantly occurs in everyday spontaneous communications, is observed as less formal and less structured. In spoken language, speakers can support their linguistic message by non-linguistic means such as facial expressions, gestures, mimics, voice qualities (volume and colour), etc. Written language is more associated with formal or significant occasions and is more formal, wordy, denser, and structured since it can be planned and organised. While spoken language typically occurs in a dialogue with another person, written language is generally more impersonal and is a monologue. Crystal (2018) considered that spoken language is time-bound, dynamic, and transient, while written language is space-bound, static, and permanent. According to Johnstone & Andrus (2024), written language is perceived as more prestigious, formal and cold than speech. It has to be edited and standardised to ensure clarity. Speech, on the contrary, is more flexible than writing because it is constantly changing.

Conclusion and future trajectories in linguistics

The present publication introduces the introductory outline of linguistics as a systematic study of language. Understandably, no book can cover the topic's whole extent and complexity, especially when the subject is as complicated as human language.

As linguistic research evolves, the coming decades promise groundbreaking advancements across multiple domains. The study of language, long regarded as a multidisciplinary field, is poised to benefit from increasingly sophisticated methodologies, interdisciplinary collaboration, and technological innovation. Future research endeavours will deepen our understanding of language structures and functions and provide practical applications that influence education, healthcare, artificial intelligence, and social policy.

With rapid progress in contemporary psycholinguistics, essential advances in artificial intelligence models simulating human-like linguistic processing and deeper insights into how emotions and social context shape language use can be expected.

With the development of neuro-imaging technology (fMRI, EEG, and MEG), neurolinguists, in cooperation with neurologists, better understand the neural mechanisms behind language processing. Current research delves into aphasia, brain plasticity, and language recovery after injury. Future developments will likely involve brain-computer interfaces (BCIs) for communication in individuals with severe speech impairments and more precise mapping of neural pathways involved in multilingualism and language disorders.

Within sociolinguistics, contemporary research focuses on new topics such as digital communication, language and gender, and the effects of migration on linguistic diversity. In the future, a greater emphasis on how social media and artificial intelligence influence language change, as well as studies on the resilience of minority languages in an increasingly globalised world, can be anticipated.

Current research in cultural linguistics highlights linguistic relativity, multilingual cultural identity, and Indigenous language preservation. In the coming years, interdisciplinary work with anthropology and cognitive science is expected to expand, further clarifying how language encodes cultural knowledge and how globalization affects linguistic diversity.

And finally, computational linguistics, which develops algorithms for language processing, powering technologies like machine translation, speech recognition, and chatbots. Current research focuses on large language models (LLMs), multilingual NLP,

and ethical AI language use. Future trends include improved contextual understanding in AI, more human-like conversational agents, and multilingual NLP systems that better handle low-resource languages, enhancing global communication and accessibility.

In sum, the future of linguistics is marked by theoretical innovation and real-world impact. As new methodologies emerge and technology continues to revolutionise data collection and analysis, linguistic research will play a central role in addressing fundamental questions about human cognition, social structures, and the nature of communication itself. These anticipated developments underscore the enduring relevance of linguistics as both a scientific discipline and a tool for understanding the complexities of human interaction in an increasingly interconnected world.

References

- Acquaviva, P. (2017). Number in Language. *Oxford Research Encyclopedia of Linguistics* <https://doi.org/10.1093/acrefore/9780199384655.013.61>
- Akinola, S. T. (2023). The ubiquity of signs. *Global Online Journal of Academic Research (GOJAR)*, 2(1), 37-49. <file:///C:/Users/Admin/Downloads/3067-6057-1-SM.pdf>
- Andrews, T. (2024). Which Language Is Richest In Words? Interpreters and Translators, Inc. <https://ititranslates.com/which-language-is-richest-in-words/>
- Appel-Meulenbroek, R. (2013). *How to measure added value of CRE and building design*. Thesis. Technische Universiteit Eindhoven.
- Austin, P. K., & Sallabank, J. (Eds.) (2011). *The Cambridge Handbook of Endangered Languages*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511975981>
- Barnard, A. (2010). Mythology and the evolution of language. In A. D. M. Smith et al. (Eds.), *The Evolution of Language: Proceedings of the 8th International Conference (EVOLANG8)* (pp. 11-18). <https://doi.org/10.1142/7624>.
- Bloch, B. & Trager, G. L. (1942). *Outline of Linguistic Analysis*. Linguistic Society of America at the Waverly Press.
- Baerman, M., & Brown, D. (2013). Case Syncretism. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Bloomfield, L. (1933). *Language*. Chicago: University of Chicago Press.
- Bradfield, J. (2014). Clicks, concurrency and Khoisan. *Phonology*, 31(1), 1–49. <https://doi.org/10.1017/S0952675714000025>
- Breedveld, J. O. (1995). *Form and Meaning in Fulfulde: A Morphophonological Study of Maasinankoo*. Leiden: Research School CNWS. <https://scholarlypublications.universiteitleiden.nl/handle/1887/68663>
- Brenzinger, M., & Shah, S. (2023). A typology of the use of clicks. *Stellenbosch Papers in Linguistics Plus*, 67, 59-77. doi: 10.5842/67-1-1007
- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511805226>
- Canepari, L. (2023). canIPA Natural Phonetics. <http://canipa.net/doku.php>
- Chandler, D., & Munday, R. (2011). *A Dictionary of Media and Communication*. Oxford: Oxford University Press.
- CheyenneLanguage.org (online). Comments on Cheyenne vowel marks. <https://www.cheyennelanguage.org/marks.htm>
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. M.I.T. Press.
- Chomsky, N. (2006). *Language and Mind*. Cambridge: Cambridge University Press.

-
- Christiansen, M.H., & Kirby, S. (Eds.). (2003). *Language Evolution: The States of the Art*. Oxford: Oxford University Press.
- Comrie, B. (2013). Writing Systems. In: M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
<https://doi.org/10.5281/zenodo.13950591>
- Connell, B. (1994). The structure of labial-velar stops. *Journal of Phonetics*, 22, 441-476.
- Cook, G., & Seidlhofer, B. (1996). *Principle and practice in applied linguistics*. Oxford University Press.
- Corbett, G. G. (2013a). Number of genders. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
<https://doi.org/10.5281/zenodo.13950591>
- Corbett, G. G. (2013b). Sex-based and non-sex-based gender systems. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Coulmas, F. (1996). *Encyclopedia of writing systems*. Oxford: Blackwell.
- Coulmas, F. (2003). *Writing Systems: An introduction to their linguistic analysis*. Cambridge University Press.
- Crystal, D. (1999). The future of Englishes. *English Today*, 15(2), 10-20.
- Crystal, D. (2010). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Crystal, D. (2018). *The Cambridge Encyclopedia of the English Language*. 3rd ed. Cambridge: Cambridge University Press.
- Crystal, D., & Robins, R. H. (2024). Language. *Britannica.com*.
<https://www.britannica.com/topic/language>
- Ćwiek, A. et al. (2021). The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377, 1841. <https://doi.org/10.1098/rstb.2020.0390>
- Cysouw, M. (2013). Inclusive/Exclusive Distinction in Independent Pronouns. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Dahl, Ö., & Velupillai, V. (2013a). The Past Tense. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
<https://doi.org/10.5281/zenodo.13950591>
- Dahl, Ö., & Velupillai, V. (2013b). The Future Tense. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
<https://doi.org/10.5281/zenodo.13950591>
- Dahl, Ö., & Velupillai, V. (2013c). Perfective/Imperfective Aspect. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
Zenodo. <https://doi.org/10.5281/zenodo.13950591>

-
- Dediu, D., Lin, J., Moisik, S. R., & Moran, S. (2023). Dental fricatives: Patterning, evolution, and factors affecting a rare class of speech sounds. In F. A. Karakostis, & G. Jäger (Eds.), *Biocultural Evolution: An Agenda for Integrative Approaches* (pp. 143–178). Tübingen: Kerns Verlag. <https://doi.org/10.51315/9783935751384.007>
- De Saussure, F. (1916). *Cours de linguistique générale*. Lausanne – Paris: Payot.
- Deverell, M. (2003). Non-human communication. In Aminoff, M. J., & Daroff, R. B. (Eds.), *Encyclopedia of the Neurological Sciences* (pp. 746–749). <https://doi.org/10.1016/B0-12-226870-9/00284-7>
- Dryer, M. S. (2013a). Coding of nominal plurality. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Dryer, M. S. (2013b). Definite Articles. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Dryer, M. S. (2013c). Order of Subject, Object and Verb. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2024). *Ethnologue: Languages of the World*. 27th edition. Dallas, Texas: SIL International. <http://www.ethnologue.com>
- Esposito, C. M. et al. (2019). Distinguishing breathy consonants and vowels in Gujarati. *Journal of South Asian Languages and Linguistics*, 6(2), 215–243. <https://doi.org/10.1515/jsall-2019-2011>
- Evans, N. (2010). *Dying Words: Endangered Languages and What They Have to Tell Us*. Malden, MA: Wiley-Blackwell. <https://doi.org/10.1002/9781444310450>
- Everett, D. Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology*, 46(4), 621–646.
- Fellbaum, Ch. (2014). Large-scale lexicography in the digital age. *International Journal of Lexicography*, 27(4), 378–395. <https://doi.org/10.1093/ijl/ecu018>
- Ferenčík, M. (2016). *English Stylistics as Discourse Analysis*. Prešov: Prešov University.
- Ferenčík, M., & Bednárová-Gibová, K. (2024). *A Handbook of Research Methods in Linguistics and Translation Studies*. Prešov: Prešov University. <http://www.pulib.sk/web/kniznica/elpub/dokument/Ferencik5>
- Gagliano, M. (2013). Green symphonies: a call for studies on acoustic communication in plants. *Behavioral Ecology*, 24(4), 789–796. <https://doi.org/10.1093/beheco/ars206>
- Gallagher, B. (2014). A conversation with Koko. <https://www.youtube.com/watch?v=SNuZ4OE6vCk>
- Ganpule, S. P. (2013). Myths of the Origin of Language in World Mythologies. *Multidisciplinary Research Journal*, 3, 188–194. <https://standrewscollege.ac.in/wp->

[content/uploads/2018/06/Myths-of-the-Origin-of-Language-in-World-Mythologies.pdf](#)

- García-Servín, M. Á., Mendoza-Sánchez, M. & Contreras-Medina, L. M. (2021). Electrical signals as an option of communication with plants: a review. *Theoretical and Experimental Plant Physiology*, 33, 125–139. <https://doi.org/10.1007/s40626-021-00203-3>
- Gasparri, L., & Marconi, D. (2024). Word Meaning. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer 2024 Edition). <https://plato.stanford.edu/archives/sum2024/entries/word-meaning/>
- Gerfen C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzacoapan Mixtec. *Journal of Phonetics*, 33(3), 311–334. DOI:[10.1016/j.wocn.2004.11.002](https://doi.org/10.1016/j.wocn.2004.11.002)
- Gil, D. (2013). Para-linguistic usages of clicks. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/142>
- Goedemans, R., & van der Hulst, H. (2013). Rhythm Types. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Gordon, G. N. (2024). Communication. *Encyclopaedia Britannica*. <https://www.britannica.com/topic/communication/Types-of-communication>
- Graz Database on Reduplication (online). <https://reduplication.uni-graz.at/redup/>
- Güldemann, T. (2007). Clicks, genetics and “proto-world” from a linguistic perspective. *University of Leipzig Papers on Africa, Languages and Literatures*, 29, 1–35.
- Hall, R. A. (1968). *An Essay on Language*. Philadelphia & New York: Chilton.
- Halliday, M. A. K. (2013). *Halliday's Introduction to Functional Grammar*. Revised by C. M. I. M. Matthiessen. Routledge.
- Hammarström, H. (2016). Linguistic diversity and language evolution. *Journal of Language Evolution*, 1(1), 19–29. <https://doi.org/10.1093/jole/lzw002>
- Heil, M. (2009). Plant Communication. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0021915
- Higgins, E., & Semin, G. (2000). Communication and social psychology. *International Encyclopedia of the Social & Behavioral Sciences* (pp. 2296–2299). <https://doi.org/10.1016/B0-08-043076-7/01816-7>
- Hobbs, M. (2023). How to Learn and Remember Chinese Tones for the Rest of Your Life. *Learn Chinese*. <https://ltl-chengdu.com/chinese-tones/>
- Hockett, C. F., & Altmann, S. A. (1968). A note on design features. In T. Sebeok (Ed.), *Animal communication: techniques of study and results of research* (pp. 61–72). Bloomington: Indiana University Press.
- Hohaus, V., & Bochnak, M. R. (2020). The grammar of degree: gradability across languages. *Annual Review of Linguistics*, 6, 235–59. <https://doi.org/10.1146/annurev-linguistics-011718-012009fhohaus>

-
- International Phonetic Alphabet (online).
<https://www.internationalphoneticalphabet.org/ipa-sounds/ipa-chart-with-sounds/>
- Johansson, S. (2005). *Origins of Language: Constraints on Hypotheses*. Amsterdam and Philadelphia, Pa.: John Benjamins.
- Johnstone, B., & Andrus, J. (2024). *Discourse Analysis*, 4th edition. Wiley-Blackwell.
- Karpinski, M. (2014). The sounds of language. In Nau, N. et al., *Book of Knowledge of Languages in Danger*. <http://languagesindanger.eu/book-of-knowledge/the-sounds-of-language/#Chapter%20contents>
- Key, M. R., & Comrie, B. (Eds.) (2023). *The Intercontinental dictionary series*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://ids.clld.org>
- Khan Academy (online). Animal communication. Available on: <https://www.khanacademy.org/science/ap-biology/ecology-ap/responses-to-the-environment/a/animal-communication>
- Kirk, P. L., Ladefoged, J., & Ladefoged, P. (1993). Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In A. Mattina & T. Montler (Eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson* (pp. 435–450). Missoula: University of Montana Press.
- Klegr, A. (2018). Language is embiggened by words that don't exist: The case of a circumfix. *Linguistica Pragensia*, 2018(1), 53-70.
- Krauss, R. (2000). Verbal Communication, Psychology of. *International Encyclopedia of the Social & Behavioral Sciences* (pp. 16161-16165). <https://doi.org/10.1016/B0-08-043076-7/01815-5>
- Ladefoged, P. (2005). *Vowels and consonants: an introduction to the sounds of languages*. Malden: Blackwell.
- Ladefoged, P. & Maddieson, I. (2008). *The Sounds of the World's Languages*. Malden, MA: Blackwell.
- Lapum, J. et al. (Eds.). (2020). *Introduction to Communication in Nursing*. <https://pressbooks.library.torontomu.ca/communicationnursing/chapter/transaction-model-of-communication/>
- Leech, G. (1981). *Semantics. The Study of Meaning*. 2nd edition. Harmondsworth: Penguin Books.
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (2014). *Ethnologue: languages of the world*. 17th ed. SIL International, Dallas, TX. <http://www.ethnologue.com>.
- Li, S., Xu, L. D., & Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, 17, 243–259. DOI 10.1007/s10796-014-9492-7.
- Lionnet, F. (2020). Paralinguistic use of clicks in Chad. In B. Sands (Ed.), *Click Consonants (Empirical Approaches to Linguistic Theory 15)* (pp. 422–437). Leiden, Boston: Brill. https://doi.org/10.1163/9789004424357_015
- Lonet Academy (2024). Learn Arabic: the Million-Word Language. Interesting facts about Arabic language. *Lonet Academy*. <https://lonet.academy/blog/learn-the-million-word-arabic-language/>

-
- Lyons, J. (1981). *Language and Linguistics*. Cambridge University Press
- Maddieson, I. (2013). Presence of Uncommon Consonants. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>. <http://wals.info/chapter/19>.
- Maddieson, I., & Precoda, K. (1989). Updating UPSID. *Journal of the Acoustical Society of America*, Suppl. 1, 86, 19. http://web.phonetik.uni-frankfurt.de/upsid_info.html
- Malchukov, A. L., & Spencer, A. (Eds.) (2012). *The Oxford Handbook of Case*. OUP. <https://doi.org/10.1093/oxfordhb/9780199206476.001.0001>
- Malden, M. A., Moran, S., & McCloy, D. (2019). PHOIBLE: Phonetics Information Base and Lexicon. cldf-datasets/phoible: PHOIBLE 2.0 as CLDF dataset (v2.0). Zenodo. <https://doi.org/10.5281/zenodo.2593234>
- Mazzaro, N. (2010). Changing perceptions: the sociophonetic motivations of the labial velar alternation in Spanish. In M. Ortega-Llebaria (Ed.), *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology* (pp. 128-145). Somerville, MA: Cascadilla Proceedings Project.
- McGregor, W. (2009). *Linguistics—an introduction*. London: Continuum IPG.
- Miller, A. (2011). The representation of clicks. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology*, Vol. 1, 416–439.
- Nau, N. (2014). Exploring linguistic diversity. In N. Nau et al., *Book of Knowledge of Languages in Danger*. <http://languagesindanger.eu/book-of-knowledge/>
- Ogasawara, N. (2005). Processing voiceless vowels in Japanese: Effects of language-specific phonological knowledge. *Acoustical Society of America Journal*, 117(4), 2460-2460. DOI: [10.1121/1.4787284](https://doi.org/10.1121/1.4787284)
- Oxford English Dictionary (online). Language. https://www.oed.com/dictionary/language_n?tl=true
- Oyeleye, A. (2004). Communication. *Encyclopedia of Social Measurement* (pp. 379-390). <https://doi.org/10.1016/B0-12-369398-5/00271-1>
- Paulin, Pascale. 2010. *Les Baka du Gabon dans une dynamique de transformations culturelles. Perspective linguistiques et anthropologiques*. Thèse du doctorat, Université Lumière Lyon 2. [available on-line at: http://www.ddl.ish-lyon.cnrs.fr/fulltext/Paulin/Paulin_2010_These.pdf]
- Pavlík, R. (2004). Slovenské hlásky a medzinárodná fonetická abeceda. *Jazykovedný časopis*, 55(2), 87-109.
- Pavlík, R. (2017). *A Textbook of English Lexicology I. Word Structure, Word-Formation, Word Meaning*. Bratislava: Z-F ingua.
- Pavlík, R. (2018). *A Textbook of English Lexicology II. Word Relations, Words in Use, Lexical Variation and Change*. Bratislava: Z-F Lingua.
- Payne, D. L., & Payne, T. (1990). Yagua. In D. C. Derbyshire, & G. K. Pullum (Eds.), *Handbook of Amazonian Languages 2*, 249-474. Berlin: Mouton de Gruyter.
- Powell, B. B. (2012). *Writing: Theory and History of the Technology of Civilization*. Wiley-Blackwell.

-
- Pretz, K. (2013). The next evolution of the Internet. *IEEE Spectrum*.
<http://theinstitute.ieee.org/technology-focus/technology-topic/the-next-evolution-of-the-internet>
- Proctor, M., Zhu, Y., Lammert, A., Toutios, A., Sands, B., & Narayanan, S. (2020). Studying clicks using real-time MRI. In B. Sands (Ed.), *Click Consonants (Empirical Approaches to Linguistic Theory 15)* (pp. 210–240). Leiden, Boston: Brill.
https://doi.org/10.1163/9789004424357_007
- Reid, N. (1997). Class and Classifier in Ngan'gityemerri. In M. Harvey, & N. Reid (Eds.), *Nominal Classification in Aboriginal Australia* (pp. 165–228). Amsterdam: John Benjamins. <https://doi.org/10.1075/slcs.37.10rei>
- Repka, R. (2020). Towards a definition of language. In *Philologia*, 30(1), 149–170.
https://www.fedu.uniba.sk/fileadmin/pdf/Sucasti/Ustavy/Ustav_filologickych_studii/Philologia/Philologia_2020-1/PHILOLOGIA_XXX_1_2020_Repka.pdf
- Robinson, A. (2009). *Writing and Script: A Very Short Introduction*. Oxford University Press.
- Roscoe, P. (2003). Mountain Arapesh. In C. R. Ember, & M. Ember (Eds.), *Encyclopedia of Sex and Gender*. Springer, Boston, MA. https://doi.org/10.1007/0-387-29907-6_69
- Rose, K., Eldridge, S., & Chapin, L. (2015). *The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World*. The Internet Society. <https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-IoT-Overview-20151221-en.pdf>
- Rubino, C. (2013). Reduplication. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online (WALS)*. Zenodo.
<https://doi.org/10.5281/zenodo.13950591>
- Sakurai, Y., Ishizaki, S. (2024). Plant–plant communication in *Camellia japonica* and *C. rusticana* via volatiles. *Scientific Reports*, 14, 6284. <https://doi.org/10.1038/s41598-024-56268-y>
- Sapir, E. (2021). *Language: An introduction to the study of speech*. New York: Harcourt, Brace & World.
- Sarkar, S. (2024). 8 Communication models: Understanding models of communication for better business interactions. <https://www.prezent.ai/zenpedia/communication-models#the-8-models-of-communication>
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Searle, J. (1992). *The Rediscovery of the Mind*. Mass, US: MIT Press.
- Searle, J. (1999). *Mind, Language and Society*. London: Orion Books.
- Shanin, K. (2011). Pharyngeals. In *The Blackwell Companion to Phonology* (pp.1–24). John Wiley & Sons. DOI:[10.1002/9781444335262.wbctcp0025](https://doi.org/10.1002/9781444335262.wbctcp0025)

-
- Shannon, C., & Weaver, W. (1948). The Mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623-656. <http://dx.doi.org/10.1002/j.1538-7305.1948.tb00917.x>
- Sittig, D. F. (2017). *Clinical Informatics Literacy: 5000 Concepts That Every Informatician Should Know*. Academic Press.
- Staes, N., Sherwood, C.C., Wright, K. et al. (2017). FOXP2 variation in great ape populations offers insight into the evolution of communication skills. *Scientific Reports*, 7, 16866 (2017). <https://doi.org/10.1038/s41598-017-16844-x>
- Štekauer, P., Valera, S., & Körtvélyessy, L. (2012). *Word-formation in the World's Languages: A Typological Survey*. Cambridge: Cambridge University Press.
- Swadesh, M. (1950). Salish Internal Relationships. *International Journal of American Linguistics*, 16, 157–167.
- The Gorilla Foundation. (online). <https://www.koko.org/communication/>
- Toki Pona (online). <https://tokipona.org/>
- UNESCO (online). *The World Atlas of Languages*. <https://en.wal.unesco.org/discover/languages>
- Vakoch, D. A. (Ed.). (2014). *Archaeology, anthropology, and interstellar communication*. NASA. https://www.nasa.gov/wp-content/uploads/2015/01/archaeology_anthropology_and_interstellar_communication_tagged.pdf
- Wacewicz, S., & Żywczyński, P. (2015). Language Evolution: Why Hockett's Design Features are a Non-Starter. *Biosemitotics*, 8, 29–46. <https://doi.org/10.1007/s12304-014-9203-2>
- Walker, W. (1972). Toward the Sound Pattern of Zuni. *International Journal of American Linguistics*, 38(4), 240-259. <https://www.jstor.org/stable/1264302>
- Wandel, A. (2015). How Many People Speak Esperanto? Esperanto on the Web. *Interdisciplinary Description of Complex Systems*, 13(2), 317-320. DOI:10.7906/indec.13.2.9
- Wicherkiewicz, T. (2014). Writing. In N. Nau et al., *Book of Knowledge of Languages in Danger*. <http://languagesindanger.eu/book-of-knowledge/writing/#ch5s1>
- Wood, G. (2022). Which Language Has the Most Words? Rosetta Stone. <https://blog.rosettastone.com/which-language-has-the-most-words/>
- Yang, A. (2023). Plants can talk. Yes, really. Here's how. National Geographic. <https://www.nationalgeographic.com/science/article/plants-can-talk-yes-really-heres-how>
- Zadka, M. (2019). Semasiographic principle in Linear B inscriptions. *Writing Systems Research*. <https://doi.org/10.1080/17586801.2019.1588835>
- Zúñiga, F., & Kittilä, S. (2019). *Grammatical voice*. Cambridge: Cambridge University Press.

Title: Basics of Linguistics for future teachers of English

Author:

prof. PaedDr. Silvia Pokrivčáková, PhD., © 2023

Reviewers:

prof. PhDr. Gabriela Lojová, PhD. (Comenius University in Bratislava)

prof. PhDr. Milan Ferenčík, PhD. (University of Prešov)

Published in 2023 by Gaudeamus, the publishing house of the University of Hradec Králové, as its 1,858th publication

1st edition (87 norm pages)

ISBN 978-80-7435-924-8



ISBN 978-80-7435-924-8

