

Essentials of Language Typology

LÍVIA KÖRTVÉLYESSY

Lívia Körtvélyessy

Essentials of Language Typology

KOŠICE 2017

© Livia Körtvélyessy, *Katedra anglistiky a amerikanistiky, Filozofická fakulta UPJŠ v Košiciach*

Recenzenti: Doc. PhDr. Edita Kominarecová, PhD.

Doc. Slávka Tomaščíková, PhD.

Elektronický vysokoškolský učebný text pre Filozofickú fakultu UPJŠ v Košiciach. Všetky práva vyhradené. Toto dielo ani jeho žiadnu časť nemožno reprodukovať, ukladať do informačných systémov alebo inak rozširovať bez súhlasu majiteľov práv.

Za odbornú a jazykovú stránku tejto publikácie zodpovedá autor. Rukopis prešiel redakčnou a jazykovou úpravou.

Jazyková úprava: Steve Pepper

Vydavateľ: Univerzita Pavla Jozefa Šafárika v Košiciach

Umiestnenie: <http://unibook.upjs.sk>

Dostupné od: február 2017

ISBN: 978-80-8152-480-6

Table of Contents

Table of Contents i

List of Figures iv

List of Tables v

List of Abbreviations vi

Preface vii

CHAPTER 1 What is language typology? 1

Tasks 10

Summary 13

CHAPTER 2 The forerunners of language typology 14

Rasmus Rask (1787 - 1832) 14

Franz Bopp (1791 – 1867) 15

Jacob Grimm (1785 - 1863) 15

A.W. Schlegel (1767 - 1845) and F. W. Schlegel (1772 - 1829) 17

Wilhelm von Humboldt (1767 – 1835) 17

August Schleicher 18

Neogrammarians (Junggrammatiker) 19

The name for a new linguistic field 20

Tasks 21

Summary 22

CHAPTER 3 Genealogical classification of languages 23

Tasks 28

Summary 32

CHAPTER 4 Phonological typology 33

Consonants and vowels 34

Syllables 36

Prosodic features 36

Tasks 38

Summary 40

CHAPTER 5 Morphological typology 41

Morphological classification of languages (holistic approach) 42

From holistic to partial approach 49

Tasks 51

Summary 56

CHAPTER 6 Syntactic typology 57

Order of sentence elements 57

Relationship between syntactic elements 59

Clause types 60

Tasks 61

Summary 66

CHAPTER 7 Lexical typology 67

The onomasiological perspective 68

The semasiological perspective 68

Lexicon-grammar borderline perspective 70

3 different foci of lexical typology 70

Natural Semantic Metalanguage (NSM) 71

Tasks 74

Summary 79

CHAPTER 8 Areal typology 80

Geographical region, neighbouring languages 81

Language families and linguistic areas 83

Common structural properties 85

Sprachbund 85

Standard Average European 89

Tasks 91

Summary 95

CHAPTER 9 Language universals 96

Quest for the right label 96

Approaches to language universals 99

Types of universals 102

Synchrony vs. diachrony, language change, explanation of universals 107

Language universals and linguistic typology 108

Tasks 109

Summary 111

CHAPTER 10 How to approach language typology? 112

Prague Linguistic School 112

Leningrad Typology Group 116

The UNITYP research group 117

The EUROTYP project 118

Canonical typology 119

Košice typology group 120

Tasks 121

Summary 124

CHAPTER 11 How do typologists work? 125

Research strategies of typological research 125

Language sampling 126

Data collection 128

Fieldwork and language documentation 130

Tasks 135

Summary 137

References 137

List of Figures

Figure 1 Family tree of the British royal family	12
Figure 2 Indo-European family tree	19
Figure 3 Uralic language family	22
Figure 4 The Indo-European language family	26
Figure 5 The Indo-European language family in the world	27
Figure 6 Classification in biology	29
Figure 7 father in the Indo-European languages	31
Figure 8 Vowels in Sentani	34
Figure 9 Vowels in Alawa	34
Figure 10 Vowels in English	35
Figure 11 Consonants in Anamuxra	38
Figure 12 Consonants in Sentani	39
Figure 13 Cross-linguistic use of word-formation processes in the research sample	55
Figure 14 Obligatory subject pronoun	65
Figure 15 Extension/denotation vs. (descriptive) meaning/sense for the temperature adjectives	69
Figure 16 The /ʊ/-/ʌ/ isogloss boundary	82
Figure 17 An isopleth sketch map of Africa based on 11 properties	83
Figure 18 Macro-areas according to Ethnologue	88
Figure 19 Areas of Europe according to Ethnologue	88
Figure 20 Definite and indefinite article	94
Figure 21 A cluster map combining nine features	94
Figure 22 #Universal 12 in The Universals Archive	103
Figure 23 Greenberg's typology of language universals	106
Figure 24 The collective method in the Leningrad Group	122

List of Tables

Table 1 Numerals in Slavic languages	3
Table 2 Numerals in Romance languages	3
Table 3 Numerals in Maltese, Hungarian and Turkish	3
Table 4 Sugar in Germanic languages	3
Table 5 Sugar in Slavic and Romance languages	4
Table 6 Sugar in Hungarian, Hebrew, Swahili and Azerbaijani	4
Table 7 Cousin in English, Slovak, Hungarian and Chinese	5
Table 8 Aspirated voiced stops became voiced fricatives and then voiced stops	16
Table 9 Voiceless stops became voiceless fricatives	16
Table 10 Voiced stops became voiceless stops	16
Table 11 The largest language families in Ethnologue	25
Table 12 Syllable structure in Comanche	39
Table 13 Slovak declension paradigm	47
Table 14 Introflection in Hebrew	48
Table 15 Index of synthesis calculated by Greenberg	53
Table 16 Examples of word order	58
Table 17 Nominal and verbal conjunction in Somali	59
Table 18 Semantic primes according to Goddard and Wierzbicka	72
Table 19 Aquamotion verbs in English and Russian	74
Table 20 Comparison of C&B verbs in English, Garifuna and Mandarin	76
Table 21 Words for pepper in Czech, Basque, Swahili and Maori	80
Table 22 SAE according to Décsy	90
Table 23 SAE according to Haspelmath	90
Table 24 SAE according to Auwera	90
Table 25 SAE according to Körtvélyessy	91
Table 26 Tea	93
Table 27 Summary of two conferences on language universals	101
Table 28 Domains in the Universals Archive	103
Table 29 Chomsky vs Greenberg and their approach to language universals	109

List of Abbreviations

ABL	ablative
ACC	accusative
AG	Agent
ART	article
DAT	dative
DISTANT.PST	distant past
FOC	focus
FUT	future
IMPF	imperfective
IND	indicative
INTRANS, INTR	intransitive
LOC	location
M	masculine
NEG	negation
O	object
PARTL	particle
PAST	past
PERF	perfektum
PL	plural
Poss	possession
PRES	present
PST.BEFORE.YEST.	past before yesterday
PTCP	participle
S	subject
SG	singular
V	verb
TNS	transitive

Preface

The 'discovery' of Sanskrit was the first important impulse for linguists (mainly in the 19th century) to seek similarities and differences among languages of the world, amplified by another crucial 'discovery', that of American Indian languages, so different from what had been believed to be a universal system of linguistic categories.

Language typology has gradually become a highly respected and widely explored field of linguistics, primarily thanks to the inspiring achievements of John Greenberg in 1960s and 1970s. Greenberg's seminal work instigated large-scale typological projects concentrated in linguistic and anthropological centres, such as, for example, *Max Planck Institute of Evolutionary Anthropology*. The significance of this development is manifested, inter alia, in the foundation of the *Association of Linguistic Typology*, organization of numerous international conferences, publication of specialized journals, such as *Linguistic Typology*, and development of large databases, for example *World Atlas of Language Structures*.

The scope of language typology as a field of research and study is enormous as it ranges over all linguistic levels like phonology, morphology, lexicon, word-formation, syntax..., and covering a range of theoretical and methodological issues.

While language typology has had a long tradition in our region dating back to the *Prague School of Linguistics*, the current situation is rather disappointing. A course of this sort is not offered by any linguistic department in either Slovakia or the Czech Republic, with the exception of the Department of British and American Studies, P.J. Šafárik University in Košice. The members of the linguistic section of this Department are very active in the field of word-formation typology both at the theoretical level and by organizing triennial conference *Typology and Universals in Word-Formation*.

This university coursebook is designed for students attending a Course into Language Typology and Universals either at the MA or PhD level. It aims to provide them with an overview of fundamental terminology, methodology, theories and key questions. This objective is pursued by the structure and contents of both theoretical sections and practical tasks.

What is language typology?

What do you know about language typology?

What does it study?

What methods can be applied to language typology?

What did Saussure understand by language?

What is langue? What is parole? What is la langue?

What is compounding? What is conversion? What is transflexion?

What paradigm/syntagm?

What is structuralism? Who are the main representatives?

What is generativism? Who were the main representatives?

The main topic of this textbook is the typological classification of languages. The introductory questions suggest that the scope of this intriguing field is very broad. Let's start from the very beginning and discuss the labels for this linguistic branch. Have a look at the following list of books written on language typology.

Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology*. Chicago: University of Chicago Press.

Haspelmath, M. et al (2001). *Language Typology and Language Universals*. Berlin: De Gruyter.

Moravcsik, E.A. (2012). *Introducing language typology*. Cambridge: CUP.

Velupillai, V. (2012). *An Introduction to Linguistic Typology*. John Benjamins Publishing Company

Song, J.J. (ed.), (2013). *The Oxford Handbook of Linguistic Typology*. Oxford: OUP

All these books deal with the typological classification of languages. However, as you may have noticed, there are two different expressions used to refer to the subject of classification – namely **language typology** and **linguistic typology**. The question, then, arises: is there a difference between them? By implication, what is the difference between language and linguistic?

One of the theoretical premises in the following chapters is the Saussurian understanding of language as a structured system of bilateral units – linguistic signs. Since Saussure we distinguish between **langue** and **langage**¹. Langage refers to the fact that we are 'pre-programmed' to acquire our mother

¹ Since both *langue* and *langage* have the same English equivalent *language*, the original terms introduced by Saussure are used here.

tongue. It is a general human ability. Langue is a system of all rules that should be observed by members of a speech community for the sake of successful communication. It is an abstract system of conventional rules that are generally recognized by speakers of a particular language, a property of society, a social phenomenon. **Parole** is an individual phenomenon, a concrete manifestation of *langue*.

Linguistic is related to the study of linguistics. Linguistics is a social science which studies human language. It is a scientific study of language.

To sum up, language refers to a system used for communication and it is the object of study in linguistics. Obviously, the difference between **linguistic** and **language** in this context is not decisive and more or less depends on the author and their theoretical background.

What seems to be of primary importance is the understanding of the word **typology**. A dictionary definition suggests that typology is “a system or the study of dividing a group of things into smaller groups according to the similar qualities they have” (Longman). In a broader sense, this definition is applicable also to language typology. **Language typology** is a system or study that divides languages into smaller groups according to similar properties they have. A crucial question concerns the shared properties which cluster languages into groups. In linguistics we usually discuss features of languages. By implication, language typology studies languages from the point of view of similar and different features. To identify similar and different features between languages, we must compare them. Thus, language (linguistic) typology is the study of similarities and differences between languages. Based on their similarities, languages can be divided into smaller groups. These smaller groups are called **language types**.

Let's discuss **language type** in a similar analytic way as we discussed language typology. Since the term **language** has been already explained, we can turn our attention to **type**. According to the Longman Dictionary of Contemporary English, type is “one member of a group of people or things that have similar features or qualities”. Synonyms are *sort, kind*. By implication, language type is characterized by specific features. Languages belong to the same language type if they share at least one characteristic, a feature typical of the particular language type. It should be noted, however, that some linguists (e.g. Croft, 2003) make a distinction between language type and linguistic type. The classification of languages into language types attempts to ‘match’ the **complete language system** with one language type. This is a **holistic approach**. In a **partial approach** the classification is based on the analysis of a selected language construction and/or phenomenon (not the entire language), for example the size of the consonantal inventory, the presence vs. absence of articles in language, the order of words in a sentence etc. This kind of grouping is called a linguistic type. Obviously, one language can belong to various linguistic types. (Croft, 2003, p. 42)

Have a look at Tables 1-2.

	1	2	3	4	5
Slovak	jeden	dva	tri	štyri	päť
Czech	jeden	dva	tři	čtyri	pět
Polish	jeden	dwa	trzy	cztery	pięć

Table 1 Numerals in Slavic languages

	1	2	3	4	5
Italian	uno	due	tre	quattro	cinque
Spanish	uno	dos	tres	cuatro	cinco
Portuguese	um	dois	três	quatro	cinco

Table 2 Numerals in Romance languages

What do you think, do these languages belong to the same type? Why? Why not?

Now study the table 3 and answer the same question as above:

	1	2	3	4	5
Maltese	wieħed	tnejn	tlieta	erba	ħamsa
Hungarian	egy	kettő	három	négy	pět
Turkish	bir	iki	üç	dört	beş

Table 3 Numerals in Maltese, Hungarian and Turkish

At first sight, the words in the first table have a lot in common. There are just some slight differences either in spelling (e.g. *dva* (in Slovak and Czech) and *dwa* (in Polish) or in pronunciation (e.g. Slovak *štyri* and Czech *čtyři*). A similar situation can be observed in Table 2. However, the languages in the last table display significant differences. Both differences and similarities entail the same question: WHY? Why do languages share some features? Why do they differ? It is language typology that attempts to find the answers.

Thus, typologists identify a feature or a set of features – e.g. words for numerals – and compare languages from the perspective of the chosen feature. They find out whether there are any similarities present or not. Moreover, it is important to identify the reason for differences. A good example of this aspect of a typologist's work is Moravcsik (2013). Our example concerns words different languages use for the concept of 'sugar'. Study the words in the table below:

English	sugar
German	Zucker
Swedish	socker
Dutch	suiker

Table 4 Sugar in Germanic languages

The form for the same concept is roughly the same in all four languages. All these languages belong to the so-called Germanic languages; it means that they have a common historical background. The formal similarities in the words representing the concept of 'sugar' can be explained by the same origin of languages. Let's have a look at the following table:

Slovak	cukor
Belorussian	цукар
Italian	zucchero
Spanish	azúcar

Table 5 Sugar in Slavic and Romance languages

Table 5 illustrates a similar situation for Slavic and Romance languages: the forms for the concept of ‘sugar’ are fairly similar. In fact, it is evident that words for ‘sugar’ are very similar in both tables, in all 3 language groups (Germanic, Slavic, Romance). In all these cases, the common origin can be a good explanation for this similarity. The Germanic languages, the Romance languages and the Slavic languages are all members of a larger group of the Indo-European language family. Thus, the words currently used for the concept of ‘sugar’ appear to be a legacy of the Proto-Indo-European, their mother language (spoken about 5000– 4000 BCE). And really, all these words come from the Sanskrit *śarkarā*. Sugarcane was first cultivated in India, the home of Sanskrit that is an Indo-European language, too. To draw a conclusion - the same or similar word-form may result from the common **origin** of languages.

If the common genetic origin were the exclusive reason for similarities between languages, we would expect languages outside the Indo-European family to have completely different words for *sugar*.

However, that is not always the case:

Hungarian	cukor
Hebrew	sukkar
Swahili	sukari
Azerbaijani	şəkər

Table 6 Sugar in Hungarian, Hebrew, Swahili and Azerbaijani

These languages are not Indo-European: Hungarian is Uralic, Azerbaijani belongs to the Altaic family, Hebrew is Afro-Asiatic, and Swahili is a Niger-Congo language. The story is that in the eighth century CE, Indian merchants began to export sugarcane; it was then brought to Europe and traders spread it around the globe along with the word itself, with spelling and pronunciation somewhat adapted to the rules of each language. Thus, similarities in word forms can result from **language contact**.

A widely discussed field of comparison in language typology is kinship terms. For example, let us compare words standing for the concept of ‘cousin’ in various languages. The criteria applied for their comparison in the table below are gender, age and paternal vs. maternal relationship:

gender	age	relationship	English	Slovak	Hungarian	Chinese
MALE	older	paternal	cousin	bratranec	unokabátya	tánggě
	older	maternal	cousin	bratranec	unokabátya	biǎogē
			cousin	-	unokafívér	-
	younger	paternal	cousin	bratranec	unokaöcs	tángdi

	younger	maternal	cousin	bratranec	unokaöcs	biǎodi
			cousin	-	unokatestvér	-
FEMALE	older	paternal	cousin	sesternica	unokanővér	tángjiě
	older	maternal	cousin	sesternica	unokanővér	biǎojiě
	younger	paternal	cousin	sesternica	unokahúg	tángmèi
	younger	maternal	cousin	sesternica	unokahúg	biǎomèi

Table 7 Cousin in English, Slovak, Hungarian and Chinese

In the English language there is just one form referring to the concept 'child of my uncle/aunt': *cousin*. Neither gender nor age are differentiated. Slovak has two different forms and the difference is based on gender – *bratranec* is a male cousin and *sesternica* a female cousin, irrespective of the age. In the Hungarian language, there are 2 criteria – gender and age. This results in four different words. In addition, Hungarian has a general word for 'male cousin' and a general word for 'cousin' (i.e., irrespective of age and gender). Chinese adds a new criterion – the paternal vs. maternal relationship. Consequently, there are eight different words. Interestingly, while English has no word for eight concepts involved in the Chinese terminology, Chinese and Slovak have no word for the general concept 'cousin'. From the point of view of the origin of languages under discussion, only English and Slovak share the same origin (Indo-European language family).

The *cousin* example shows that there are similarities and differences between languages that share the same origin as well as between genetically unrelated languages. Kinship relationships reflect the social organization of the family and/or the clan. They are an example of social and cultural environment. Consequently, similarities and differences between two **communities and cultures** are reflected in similarities and differences between two languages.

To sum it up, reasons for similarities between languages can be threefold: common historical origin, language contact, and shared language environment (natural and/or cultural). By implication, differences can be due to the exact opposite: different origin, no language contact and different language environment. Linguistic typology deals with these three possibilities and classifies languages into **types**. This classification is based on the presence vs. absence of a particular feature in a language. As Moravcsik (2013) puts it, it can be compared to the role of butter in various cuisines. In English cuisine it is used for cooking, in Tibetan cuisine it is put into tea. It is not used in Chinese cuisine at all. Similarly, a particular language feature can be present in two languages and have the same function, or it can be present in both languages but with a different function, or it is present only in one of the compared languages.

As has been mentioned, languages are similar because they share some language features or, the other way round, the differences between languages are caused by absence of some features. If they share examined features, they belong to the same language type. This situation can be compared to the system used by mobile phones. Nokia, for example, is typical of its ringing tone and the operating

system Windows. Sony Ericsson, on the other hand, uses Android. From the point of view of the operating systems they belong to different types. However, there are features Nokia and Sony share – you can carry the phones with you, they are mobile, and we use them for sending messages and phoning. Moreover, these features are typical not only of Sony and Nokia phones but of all mobile phones – they are universally shared. In analogy, language typology deals not only with typologically shared features but also with universally shared features – features that are shared by all natural languages in the world, i.e. **language universals**.

Thus, a typologist chooses a language feature and compares languages from the perspective of that particular feature. A precondition for any comparison in typological research is the existence of **tertium comparationis**, that means, a specific language feature or a group of features. A question is, however, what are those language features that can be identified in languages all over the world. Let us illustrate this point with some examples. It could be stated, for instance, that compounding is such a feature. However, if Slovak and English are to be compared from this point of view, the comparison must start with a question about the nature and the delimitation of compounding in these two languages. While *silver ring* or *language teacher* are usually classified as compounds, their Slovak equivalents *strieborný prsteň*, *učiteľ jazyka* are unambiguously analysed as phrases. While all Slovak compounds represent one orthographic unit English compounds can be spelled separately. Similar differences abound. In a similar vein, conversion as a feature raises considerable problems. An initial answer to the question about the presence of conversion in the Slovak language is negative. But this is not quite true: Slavic languages and some other inflectional languages have an analogical word-formation process called paradigmatic derivation (transflexion), i.e. the formation of a new word without any derivational suffix. Instead the whole inflectional paradigm is changed to form a new word of a different word class, for example:

Slovak: *dobrý* ‘good’ → *dobr-o* ‘the good’

Both conversion and transflexion represent non-affixal procedures of coining new complex words. (Ološtiak & Ivanová, 2016)

Another example of a language feature which can be subject to comparison is the category of case. However, case is not unproblematic either. For example, Jakobson (Jakobson, 1984) maintains that the Russian dative differs from the German dative due to different paradigmatic oppositions. He also refers to a talk of a German linguist Max Deutschbein delivered at the International Congress of Linguists in Rome in 1933. Deutschbein discussed the meaning of the system of cases in Indo-European languages and came to a conclusion that the general meaning of each case is “determined by the entire case system of a given language,” and therefore can only be established by investigating the structure of this system (Jakobson, 1984, p. 62). A similar observation can be found in the introduction to the *The*

Handbook of American Indian Languages by Franz Boas (Boas, 1911). When describing the methods of language description, he states that “[n]o attempt has been made to compare the forms of the Indian grammars with the grammars of English, Latin or even among themselves; but in each case the psychological groupings which are given depend entirely upon the **inner form of each language**” (Boas, 1911, p. 77).

Our three examples might suggest that nothing like typological comparison of languages is possible because of the non-existence of categories that could be compared. However, both Deutschbein and Boas doubt the existence of pre-established (a priori) categories. Haspelmath calls their approach **non-aprioristic structuralism** (Haspelmath, 2007).

Obviously, language typology presupposes the existence of some features/categories that can be compared and that are established in natural languages. Haspelmath offers two answers to the question of the existence of pre-established categories.

The first answer is given by the generative approach to language. While generativists do not seem to have paid much attention to this issue, they nevertheless came up with a theory of innate categories. The task of a linguist is to discover these categories. There is a set of universally available crosslinguistic categories (such as adjective, passive voice, accusative case, future tense, subject, affix etc.) and languages choose from them. The main task of typologist is to find out what crosslinguistic categories are present in analysed languages. According to Chomsky “[w]e require that the grammar of a given language be constituted in accord with a specific theory of linguistic structure in which such terms as "phoneme" and "phrase" are defined independently of any particular language” (Chomsky, 1957, p. 50). Haspelmath calls this approach **aprioristic generativism**. However, if the existence of innate, universal categories is denied, the question of the possibility of a non-aprioristic typology must be raised again. Haspelmath’s (2010) suggestion is the application of cross-linguistic concepts instead of descriptive categories (e.g. tense, affix, subject, dative etc.). **Comparative concepts** are defined by typologists and they enable them to “capture interesting generalizations ... Thus, comparative concepts cannot be right or wrong, at least not in the same sense that innate crosslinguistic categories can be right or wrong. They can only be more or less productive, in that they allow the formulation of more or less interesting subdivisions and generalizations” (Haspelmath, 2010, p. 678). He exemplifies this kind of comparative concept by the adjective:

GENERALIZATION: If a language has dominant SOV word order and the genitive follows the governing noun, then the adjective likewise follows the noun. (Greenberg 1963, #Universal 5)

DEFINITION: An adjective is a lexeme that denotes a descriptive property and that can be used to narrow the reference of a noun.

This definition makes use of the comparative concept ‘lexeme’ and the conceptual semantic concepts ‘property’ and ‘narrow the reference’ (the latter is necessary to exclude words like size and beauty, which denote properties, but are not normally used to modify a noun). Importantly, it is irrelevant for this definition whether a language has a separate word class that would be called ‘Adjective’ (i.e. a descriptive category), or whether it uses its ‘Noun’ or ‘Verb’ categories to attribute properties to nominal referents. In the latter case, a Verb (= descriptive category) may be an adjective (= comparative concept) for the purposes of crosslinguistic comparison (cf. Dryer, 2005a, p.354). Since descriptive categories and comparative concepts are different kinds of entities, there is no contradiction here².

He calls this third approach (in opposition to non-aprioristic structuralism and aprioristic generativism) **aprioristic typology**. It should be, however, noted that this kind of approach was applied before Haspelmath, for example, by Greenberg (1963) who, within his word-order typology, insisted on using semantic criteria for the definition of subject and object (cf. Chapter 6).

The discussion of comparative concepts would not be complete without reference to **basic linguistic theory**. The starting point of this theory differs from Haspelmath’s approach. While Haspelmath develops his theory for typological comparison of many languages, basic linguistic theory “consists in study and comparison of the grammatical patterns of individual languages” (Dixon, 2010, p. 5). Dixon’s three volumes are used especially by linguists who describe a single language and collect their data in the field with the aim to write a grammar of a language. According to Dixon, “no two languages are precisely the same, in any feature. Although the same labels are used for describing grammatical categories in different languages (if they were not, there would be no science of linguistics) they have

² For illustration, the English noun *beauty* expresses Quality, which is a typical feature of adjectives and may be used both attributively and predicatively like adjectives. ... It should be noted that in many languages the difference between adjectives and verbs is vague or non-existent.

Haspelmath’s proposal concerns the general problem of word-classes. On the one hand, what is called, for example, Adjective behaves differently in different languages – suffice to compare an inflectional language with rich inflectional paradigms of adjectives (e.g. most Slavic languages) and an isolating language (e.g. English) or an agglutinating language (e.g. Hungarian) with no modification of the basic form of an adjective. Moreover, there are a number of languages in which there is no separate class of adjectives, or the boundaries between adjectives and verbs / nouns are fuzzy and difficult to define. Rijkhoff (2000, p. 224), for example, gives arguments that Chinese does not have a separate class of adjectives; instead he speaks of adjectival verbs. The problem of unambiguous delimitation of word-classes is omnipresent, with diversity of its manifestations. In some languages, like Samoan, there is only one word-class, combining the functions of nouns, verbs, and adjectives. Thus, it is a common case in Indo-European languages that the boundary between verbal participles (both present and past) and adjectives is vague, which is reflected in contradictory views of this problem. Another example is categorical transition in English pointed out by Mathesius (1975, pp. 57-58). He demonstrates that adjectives characteristically expressing a permanent quality can behave like verbs that typically represent dynamic actions (e.g., He is short of money where the adjective expresses a changing, momentary characteristic of its subject rather than a permanent quality). Japanese as well as Bantu languages allow attributive finite verbs, i.e., verbs fulfilling a typical adjectival function. Moreover, Japanese, distinguishes between adjectives and nominal adjectives. Similar and many other instances of word-class fuzziness abound in languages of the world.

a slightly different signification for each language” (2010, p. 9). Dixon also claims that the main aim of basic linguistic theory is to compare similar phenomena between languages and, for this purpose, the same label to similar phenomena in different languages should be used. Any language description consists of two parts: grammar and lexicon. While grammar focuses on the underlying categories and structure, including also phonology, the lexicon deals with lexical forms. Dixon understands grammar and lexicon as separate components which, unfortunately, are studied by separate groups of scholars who apply different approaches and methodology to their analysis. He suggests a joint study of both grammar and lexicon since they are interrelated concepts (Dixon, 2010). A good illustration of this kind of approach is his description of adjectives which can be compared to Haspelmath’s category above. First, he starts with a detailed description of adjectives from the point of view of grammar:

- a) *Adjectives have similar properties to verbs. That is, an adjective can occur as head of a phrase filling predicate slot in clause structure, just as an intransitive verb may. For each such language, some criteria can be discerned to distinguish Adjective and Verb. The actual criterial properties vary from language to language; they typically include slightly different possibilities between verb and adjective for being modified when functioning as predicate, for functioning as modifier within an NP, for occurring in comparative constructions, and for forming adverbs (cf. footnote 2).*
- b) *Adjectives have similar properties to nouns. That is, an adjective may occur as modifier in an NP and it may also make up a complete NP (a decision then has to be made between saying that the adjective is NP head, and saying that a noun head has been ellipsed). An adjective may undergo the same morphological processes as a noun; for example, taking number and/or case marking. However, there always are some criteria which enable the linguist to distinguish two word classes. They may relate to the internal structure of NPs (if an adjective is head there may be fewer possibilities for modification than if a noun is head) or to the fact that only an adjective may occur in a comparative construction, or may form adverbs.*
- c) *Adjectives share grammatical properties with both verbs and nouns. An adjective can function similarly to an intransitive verb as being head of a predicate, and it may inflect like a noun when occurring in an NP.*
- d) *Adjectives have grammatical properties different from those of verbs and of nouns. English is of this type—an adjective may neither function as predicate head nor as NP head; it does not share any inflection with verb or with noun. Unlike nouns and verbs, an adjective occurs in a comparative construction (marked by either -er or more), and adverbs may be formed from many—but not all—adjectives.*

Afterwards, the description of two basic semantic tasks of adjectives follows:

- (I) *Make a statement that something has a certain property. In languages of type (a), this is achieved through the adjective functioning as intransitive predicate (literally, ‘The man happies’). In languages of types (b) and (d), the adjective will make up a copula complement or verbless clause complement argument (‘The man (is) happy’). Languages of type (c) are likely to show both possibilities.*
- (II) *Provide a specification that helps focus on the referent of the head noun in an NP. This is shown by an adjective acting as noun modifier in a language of types (b), (c), and (d), as in ‘The happy man’. Languages of type (a) differ in how adjectives fulfil this task; in some, an adjective must occur within a relative clause construction (literally, ‘The man who happies’). (Dixon, 2010, pp. 113,114)*

Classes of nouns and verbs are large and open in every language. In the case of adjectives, languages can differ – there are languages with large and open class of adjectives and languages in which the class of adjectives is small and closed. The semantic basis of a small adjective class (i.e., the language has a small number of adjectives) usually consist of the following semantic types: DIMENSION (big, little), AGE (old, young), COLOUR (black, white), VALUE (good, bad). If the class of adjectives is larger, one can expect also semantic types of PHYSICAL PROPERTY (raw, heavy). Larger classes will cover HUMAN PROPENSITY (clever, rude).

The discussion of comparative concepts also illustrates two preliminary factors that influence any research in language typology before it has even started. First, there is the theoretical background of the typologist. To put it simply, if the typologist is, for example, a structuralist, the methodology of research will differ from the generativist's work. Second, there is the aim of the work itself. There are linguists who focus on a description of a specific language (in many cases these are endangered languages). They have their own tools and methodologies. On the other hand, there are typologists who compare various languages and who could not work without descriptive grammars written by the first group.

Tasks

- 1. Typologists make use of various databases available on the Internet. Go to the following links and describe them. What is their purpose? What is their content? Who are the authors?**

<http://www.linguistic-typology.org/>

<http://www.wals.info>

<http://www.surrey.ac.uk>

<http://typo.uni-konstanz.de/archive/intro>

<http://www.ethnologue.com>

<http://reduplication.uni-graz.at/db.html>

- 2. Go to the WALS and find the chapters on Consonant Inventories, Number of Genders, Order of Degree Word and Adjective, and Tea. Answer the following questions for each of these chapters:**

- a) What field of linguistics is described in the chapter?
- b) What is the basic research question?

- c) What methodology is applied? What is the tertium comparationis? Does the typologist work with the analysed features in accordance with Haspelmath's comparative concepts?
- d) How many languages are compared?
- e) What are the most important results?

3. **Read the following paragraph** (Haspelmath, Pre-established categories don't exist - consequences for language description and typology, 2007):

Are categories totally different across languages?

Of course, there are many similarities between categories across languages, and this fact often leads to the temptation of equating language-particular categories with each other. The English Passive and the Japanese Passive share many properties, and Russian Suffixes are similar to Arabic Suffixes in many ways. These similarities are hardly accidental, and it is an important task of linguistics to find out how far they go and how they might be explained. However, it is important to realize that similarities do not imply identity: It is very hard to find categories that have fully identical properties in two languages, unless these languages are very closely related. Crosslinguistic similarities of categories are often best expressed in the form of implicational scales or semantic maps (see Croft 2001, Haspelmath 2003). In order to find generalizations of this sort, one has to start with the awareness that each language may have totally new categories.

Explain: Why should each typologist understand that each language may have totally new categories? Is the presence (or absence) of a specific feature a yes/no question?

4. **Use the English/Chinese words from Table 7 to describe the relationship between William and Peter Phillips, William and Zara Phillips. The Princess of Wales, Diana, had a brother, Charles Spencer, who has three daughters and one son. What is their relationship to the Duke of Cambridge?**

The Hungarian words for *cousin* are compounds and the first element (*unoka*) means 'grandchild'. The second element refers to a sibling – female, male, younger, older. Use these words to describe the relationships between the grandchildren of Queen Elizabeth II and Philip, Duke of Edinburgh.

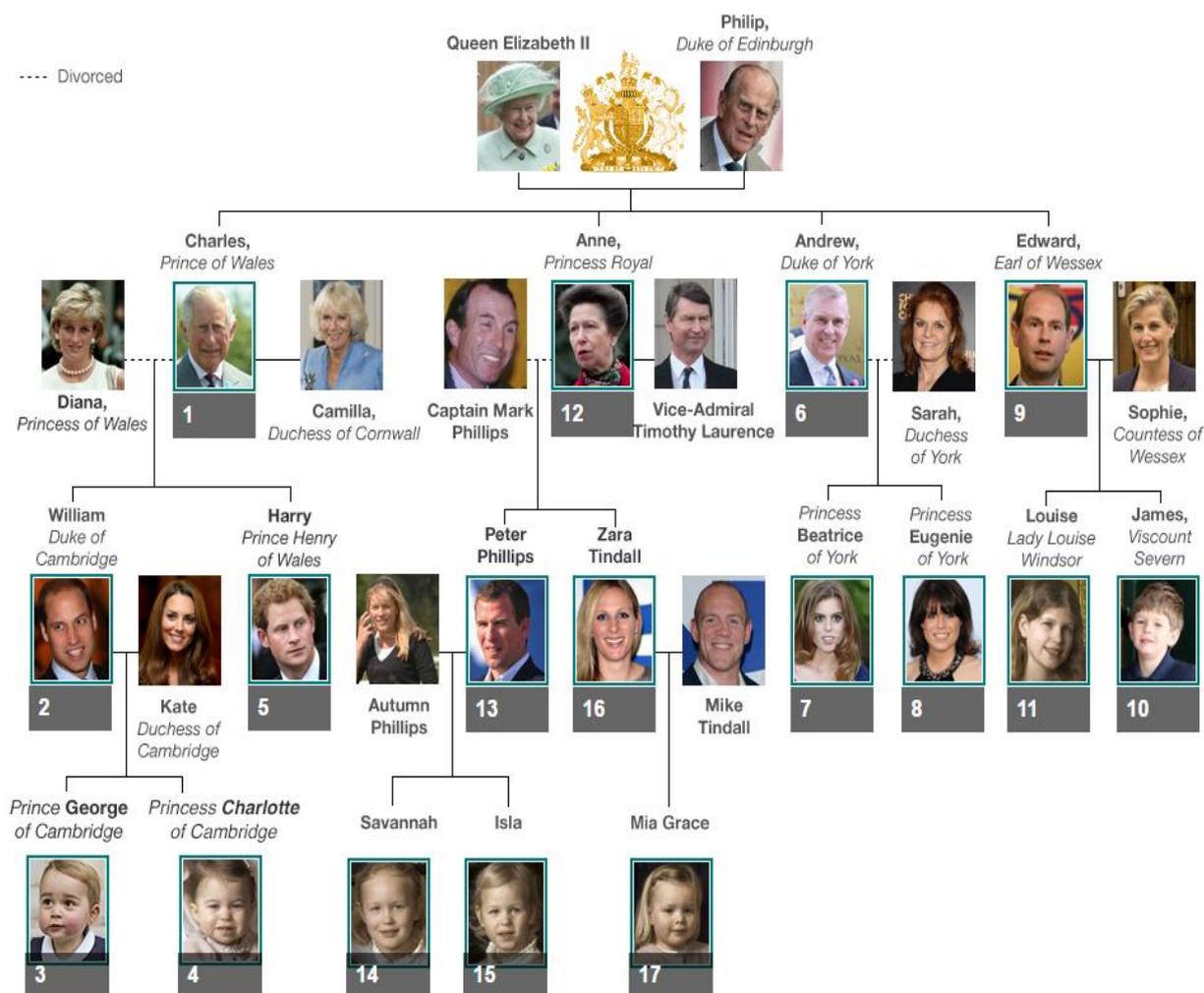


Figure 1 Family tree of the British royal family

Source: <http://www.bbc.com/news/uk-23272491>

5. Taking into consideration Haspelmath's approach to comparative concepts, do you think Dokulil's onomasiological approach and his conceptual categories could be applied as a methodology in typological research?

6. Compare three definitions of language typology below. What do they have in common? In what details do they differ? Explain why classification of languages in Croft's view is impossible?

Typology is the study of linguistic systems and recurring patterns of linguistic systems. Typology concerns itself with the difference and similarities between linguistic systems, both within and across languages, and investigates patterns of distributions of linguistic structures, something which demands cross-linguistic surveys. (Velupillai, 2012, pp. 15-36)

Traditionally, typology was used as an alternative method in pursuing one of the same goals as generative grammar: to determine the limits of possible human languages and, thereby, to contribute to a universal theory of grammar. (Bickel, 2007, p. 1)

Typology, in the sense of a cross-linguistic classification of types, is often considered to be a classification of whole languages into types, that is, a classification of **language types**. For example, Mam is an indexing genitive language type, Bulgarian is an adpositional genitive language type, etc. The notion of language type originates in the nineteenth-century view of the morphological typology of languages. In the vast majority of cases, however, a classification of language types is difficult, if not impossible. (Croft, 2003, p. 42)

- 7. Read the following paragraph from work by Dixon (Dixon, 2010, p. 9) and compare it to what was said about Jakobson's approach:**

*The label 'accusative' is typically used for an affix marking that a noun phrase (NP) is in direct object (O) function. In Quechua (spoken in the Andes), an accusative suffix goes onto the last word of the NP, but in Latin it goes onto every word in the NP. In Turkish an NP is marked by the accusative suffix only when it has definite reference. Beyond this central—and defining—function of marking O function, an accusative affix may have additional roles in the grammar of an individual language. In Latin, for example, accusative may also mark a length of time—as in *tōtam noctem dormīuī* ('all:ACCUSATIVE night:ACCUSATIVE I:slept') 'I slept the whole night'—and accusative is required after a number of prepositions, including *circum* 'around'.*

- 8. Go to the WALS database and identify cases of four different language (linguistic) types in Slovak and English.**

Summary

What is language typology?
What features are crucial for language typology?
What are comparative concepts?
What is a holistic approach to language typology?
What is a partial approach to language typology?

The forerunners of language typology

What do you know about Comparative and Contrastive linguistics?

What is Sanskrit?

What kind of language is Latin? And Greek? What do these languages have in common?

What are affixes?

What is introflexion?

What is typical of inflective languages?

The main aim of this chapter is to introduce various linguists who, with their work, laid the foundations of present-day linguistic typology. Due to its diachronic nature, this chapter is also a short discourse into the history of linguistics. However, we will not go as far as antiquity. Instead, we will start in the late 18th century when a judge named William Jones was sent to Calcutta in India. William Jones arrived in India in September 1783 to start his career at the Supreme Court. He was deeply interested in various sciences and arts and was impressed by the richness India could offer in this respect. He was one of the founders of the Asiatic Society (1784). Members of the Society met regularly and at one of the meetings William Jones delivered a speech in which he claimed that:

The *Sanskrit* language, whatever be its antiquity, is of a wonderful structure; more perfect than the *Greek*, more copious than the *Latin*, and more exquisitely refined than either, yet bearing to both of them a stronger **affinity**, both in the **roots of verbs** and the **forms of grammar**, than could possibly have been produced by accident; so strong indeed, that no philologist could examine them all three, without believing them to have sprung from some **common source** ... (Jones, 1786).

His revolutionary observation about the similarities between Greek, Sanskrit and Latin gave rise to new branches of language study – comparative and contrastive linguistics. The following sections of this chapter provides a summary of the most important contributors to the field.

Rasmus Rask (1787 - 1832)

Rasmus Rask came from Denmark. Influenced by the Romantic period he studied old Scandinavian languages and wrote a book about the origin of Old Norwegian and Icelandic.

He passed away at a relatively young age and since he wrote in Danish, many of his ideas remained unknown to the international linguistic community for many decades. However, he was the first who pointed out a number of typological issues. He drew attention to the genetic relationships between

Scandinavian and Germanic languages, Greek, Latin, Lithuanian, Armenian and Slavic languages. Later on, he completed this list with Sanskrit, Persian, Albanian and the Celtic languages. He did not follow the linguistic mainstream of his day and rejected the then fashionable idea of searching for the original language. He believed Greek to be the oldest language still in use, but – as he stressed – the language from which it had developed was unknown. Rask also paid attention to grammatical similarities between languages. He claimed that they are the most important criterion for the identification of genetic relationships between languages. In his view, lexical similarities are less reliable because they may result from language contacts. On the other hand, overlap of two languages in core vocabulary is a sign of their genetic relatedness. Rask is generally considered as one of the founders of the so-called genetic-comparative method. He created a new basis for comparative linguistics by investigating not only the languages' vocabularies, but also their phonetical and grammatical idiosyncrasies. Still, there are others from the same period who independently of Rask were able to come up with similar conclusions and systemize them into a more complex work.

Franz Bopp (1791 – 1867)

Franz Bopp, a German philologist, is considered to be the father of the genetic-comparative method. He studied Eastern languages in Paris and was the founder of the Sanskrit department at Berlin university. Bopp was influenced by Romanticism. In this movement, the main aim of philosophers, politicians, and scholars was to prove the great history of the nation that could be compared to that of French classicism and ancient Greece. This interest also concerned the history of the German language. Bopp's aim was to reconstruct the initial stage of language. However, from the present-day perspective his methodological approach to this task was more important than the aim itself: he namely compared many different languages in terms of their conjugation and declension. In his work *Vergleichende Grammatik* (Comparative grammar) he specifies three basic tasks for comparative research: to give a description of the original grammatical structure of the languages as deduced from their comparison; to trace their phonetic laws; to investigate the origin of their grammatical forms. By this comparison he established the foundations of the genetic-comparative method. Importantly, he maintains that cognate languages serve to elucidate grammatical forms lost in Sanskrit.

Jacob Grimm (1785 - 1863)

Jacob Grimm is one of the founders of the so-called historical grammar. In his famous book *Deutsche Grammatik* (1819) he formulated a law about the change of Indo-European consonants in Germanic languages. While Bopp focused on morphological changes, Grimm dealt with phonetics, and identified regularities in the changes of consonants. He introduced the idea of sound shift – *Lautverschiebung*.

Grimm described two consonant shifts that involve nine consonants. One shift (probably a few centuries before the Christian era) affected the Indo-European (IE) consonants and is evident in English, Dutch, other Low German languages, and Old Norse. The other shift (about the 6th century AD) was less radical in scope and affected the Germanic consonants, resulting in the consonant system evident in Old High German and its descendants, Middle High German and Modern High German (Standard German). According to this law, the ancient unvoiced *p, t, k* became the English unvoiced *f, th, h* and the Old High German *f, d, h*. The law further states that the ancient voiced *b, d, g* became the English unvoiced *p, t, k*, and the Old High German spirant stops *f, ts, kh*; Moreover, the originally voiced *bh, dh, gh* became the English voiced *b, d, g* and the Old High German *p, t, k*. The following tables summarize the shifts and complete them with examples (Fischer, 2006, pp. 22-23):

IE >	Proto-Germanic >	in most Germanic dialects	Example
b ^h >	ḃ >	b	Sanskrit <i>bhárāmi</i> – Old English <i>beran</i> – Modern English <i>to bear</i>
d ^h >	ḏ >	d	Sanskrit <i>bándhanam</i> (<i>bandage</i>) – Old English <i>bindan</i> – Modern English <i>to bind</i>
g ^h >	ǵ >	g	PIE <i>*longh-</i> – Old English <i>lang</i> – Modern English <i>long</i>

Table 8 Aspirated voiced stops became voiced fricatives and then voiced stops

IE > Germanic	example
p > f	Latin <i>pedem</i> ↔ English <i>foot</i>
t > θ	Latin <i>tres</i> ↔ English <i>three</i>
k > χ	Latin <i>cor</i> ↔ English <i>heart</i>

Table 9 Voiceless stops became voiceless fricatives

IE > Germanic	example
b > p	Russian <i>jabloko</i> ↔ English <i>apple</i>
d > t	Latin <i>dentis</i> ↔ English <i>tooth</i>
g > k	Latin <i>grānum</i> ↔ English <i>corn</i>

Table 10 Voiced stops became voiceless stops

The *Lautverschiebung* does not apply, for example, to loanwords that entered into the English language after the period when the shift occurred. The evidence is words like *pedestrian, tenuous, canal*.

Grimm's Law proved the genetic relatedness of the Indo-European languages by comparison of their phonetic make-up. Indisputably, his work is not only an example of historical linguistics but also of typological comparison of languages.

A.W. Schlegel (1767 - 1845) and F. W. Schlegel (1772 - 1829)

Friedrich Wilhelm von Schlegel distinguished between languages that express relations between words by inflection (internal modification) and languages that use affixes for the same purpose. Even though this typology is very simple, it states explicitly that languages of the world should be classified according to morphological criteria. His older brother, August von Schlegel extended this typology with languages which do not have inflectional morphology³. By implication, languages can be divided into three classes: (a) languages without any grammatical structure, like Chinese; (b) languages with agglutinated affixes, like Turkish; (c) languages with inflections. The last group can be further divided into synthetic languages, e.g. Latin and Ancient Greek, and analytic languages, e.g. French.

Wilhelm von Humboldt (1767 – 1835)

Philologist, scholar, diplomat, a friend of Schiller and Goethe, Wilhelm von Humboldt was also interested in languages. He studied many of them – besides Greek and Latin also Hungarian, Tatar, Semitic languages, Japanese etc. His primary aim was to found comparative anthropology. From this point of view, comparison of languages was just a tool to achieve the aim. He focused on the structure of languages. To refer to the structure, he used the term **Sprachbau** or **Organism**. The structure of languages was the starting point for their classification. He understood language as an inborn human ability that arose together with the human being as a product of his brain activity. The initial language was ideal and perfect; however, all its subsequent development meant its decline. By implication, the most perfect language is the oldest known language – Sanskrit. Within the framework of German romanticism, Humboldt claimed that language forms thinking and reflects the national soul (*Geist*). Languages differ because nations differ, they view the world from different angles. The more perfect this view, the more perfect the language. Language structures reflect the mentality of a nation. Unfortunately, these ideas were later misused, e.g. by Nazism.

What distinguished Humboldt from his contemporaries was his interest in existing languages, not only in Latin and Sanskrit. However, his methods bordered on speculations. He adopted the Schlegels' typology and extended it with the incorporating type of language. He paid attention not only to the grammatical structure of words but also to sentences: "Far more, however, than in individual words, the intellectual diversity of nations is exhibited in the constructions of speech, in the range that they are able to give to sentences" (1988:168). Humboldt was thrilled with the idea of a perfect, ideal language. He postulated three possible forms (apart from Chinese⁴) for "the attainment of sentence-

³ For more information on morphological classification of languages cf. chapter 5.

⁴ According to Humboldt (1988:216) Chinese "dispenses with all grammatical forms" – there are no grammatical forms and word order is of great significance.

making, the *inflectional*, the *agglutinating* and the *incorporative* (1988:216)⁵. Humboldt maintained that the presence of one property indicates the presence/absence of some other. Importantly, he was aware of the fact that one language can exhibit properties of various language types.

August Schleicher

August Schleicher was originally a botanist and this background led him to comparison of a natural language to an organism. Influenced by Darwin's theory of evolution, he divided the life of a language into two stages. In the first, prehistoric period, a language is born and it develops – first it is an isolating language (e.g. Chinese) than agglutinating (e.g. Hungarian) and finally, the most perfect stage in the life of a language is the inflectional period. The second, historical period, is typical of language decline that ends with the death of a language. In harmony with this approach Schleicher developed the so-called **Stammbaumtheorie**, which leans upon the similarity between genealogical relationships among languages and genealogical relationships in a human family. Based on this resemblance the language genealogical relationships can be presented in the form of a tree:

⁵ Humboldt included in his studies also languages spoken on the American continent. Nowadays the term polysynthetic is preferred.

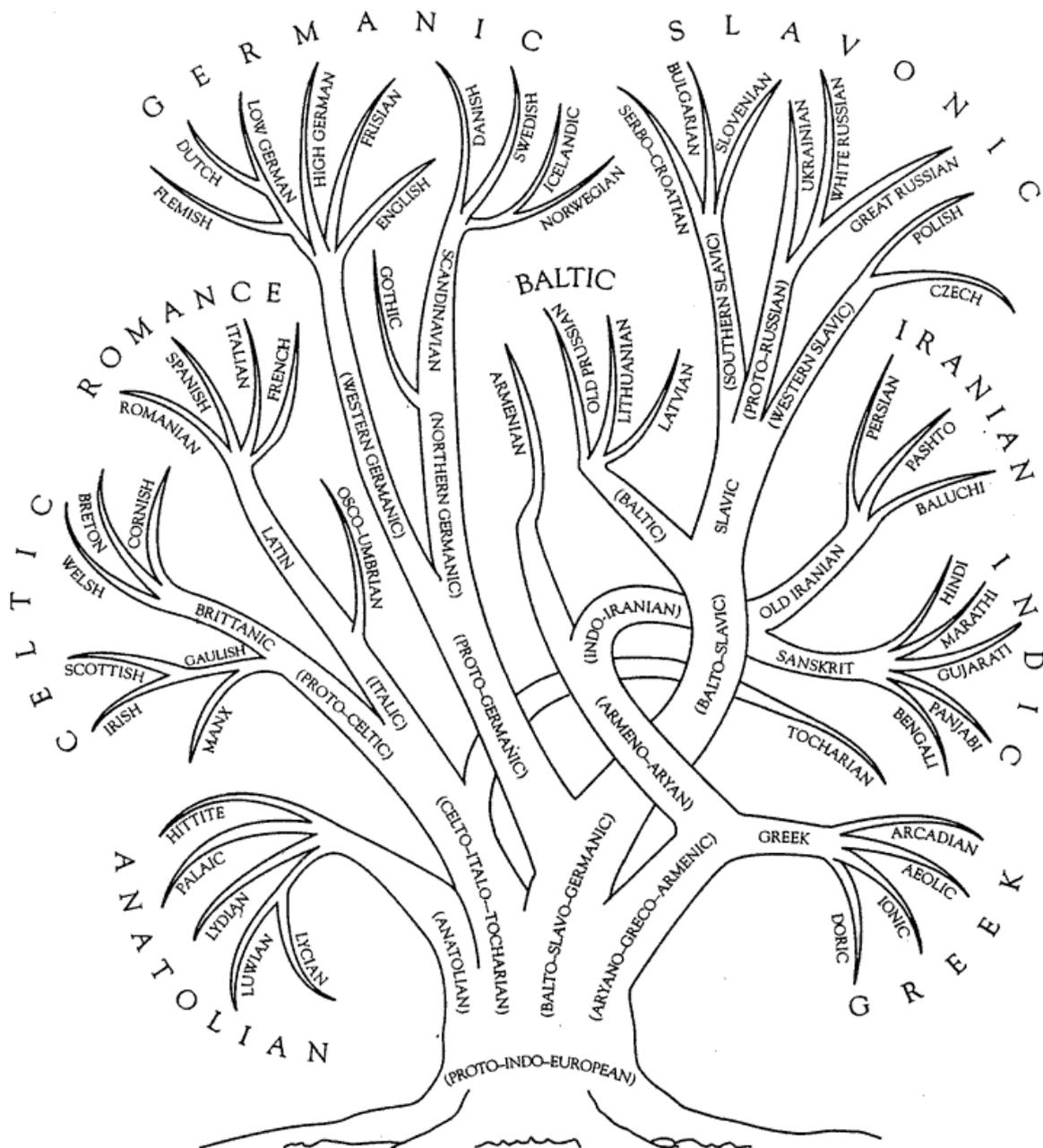


Figure 2 Indo-European family tree

source: <https://edlauber.files.wordpress.com/2013/12/schleichers-language-tree.png>

Neogrammarians (Junggrammatiker)

Junggrammatikers is a name for a movement that started as a revolt of a group of young linguists against their teachers. The movement started at the University of Leipzig and the main representatives were Karl Brugmann, Hermann Osthoff, Hermann Paul, Eduard Sievers, and Karl Verner. Their main focus was on phonetic changes from a historical perspective. The underlying idea was the regularity hypothesis. According to this hypothesis, sound changes affect each word without exception. In accordance with this approach Karl Verner completed Grimm's Law and formulated the so-called

Verner's Law that accounts for what was originally believed to be exceptions to Grimm's Law as regular changes.

The name for a new linguistic field

The period of historical and comparative linguistics developed its own methodology that became a starting point for language typology. While it concentrated on language development, attention was primarily paid to phonetic changes and the morphological structure of languages. This led to comparison of languages and to the identification of their affinities. This period engendered initial classifications of languages, and indicated two main perspectives of research in this field. First, the historical perspective was aimed at the classification of languages from the genealogical point of view. Second, the structural perspective attempted to classify languages on the basis of their structure.

The climax of this period was reached by Georg von Gabelentz (1840 - 1893) who claimed: "If one had to baptize a not yet born child, I would choose the name **typology**. I see here a task for general linguistics, whose solution can already be tempted with the means we have now at our disposal" (quoted from (Sgall P. , 1995, p. 10). He thus was the first to use the word *typology*. The language typology based on grammatical forms and structure was further developed by Edward Sapir (1884-1939) although he refers to classification of languages as a "fruitless undertaking" (Sapir, 2005, p. 130). Sapir suggests a classification according to the formal processes most typically developed in individual languages and divides languages into four types: isolating, prefixing, suffixing and symbolic languages: *"Those languages that always identify the word with the radical element would be set off as an "isolating" group against such as either affix modifying elements (affixing languages) or possess the power to change the significance of the radical element by internal changes (reduplication; vocalic and consonantal change; changes in quantity, stress, and pitch). The latter type might be not inaptly termed "symbolic" languages. The affixing languages would naturally subdivide themselves into such as are prevailingly prefixing, like Bantu or Tlingit, and such as are mainly or entirely suffixing, like Eskimo or Algonkin or Latin"* (2005, p.133).

In addition, he developed a conceptual classification and divided languages into pure-relational and mixed-relational, both groups further divided into simple and complex subgroups.⁶

What all these classifications have in common is the effort to classify languages as a whole. Since they are concerned with wholes rather than their parts this approach is also called **holistic**.⁷ Sapir, for example, states that "[I]t must be obvious to anyone who has thought about the question [of the general form of a language] or who has felt something of the spirit of a foreign language that there is

⁶ For more information on this topic cf. Chapter 5 on morphological typology of languages.

⁷ Cf. also Chapter 1 on holistic vs. partial approach to language typology.

such a thing as a basic plan, a certain cut, to each language. This type or plan or structural 'genius' of the language is something much more fundamental. Much more pervasive than any single feature of it that we can mention, nor can we gain an adequate idea of its nature by a mere recital of the sundry facts that make up the grammar of the language ' (Sapir, 1921, p. 20). This approach was also applied by the Prague linguistic school and their **characterology**.⁸ However, the next development of the field shifted the focus to various specific areas of language, e.g. word-order, relative clauses, passives, articles. In opposition to the holistic approach, this method is called the **partial** approach.

The main aim of this chapter was to introduce the forerunners of the present-day typology and to give an overview of the historical development of the field. For the description of the present day approaches and methods in language typology refer to chapter 10.

Tasks

1. **Read the following quote from Rasmus Rask (1818)** (the translation by Pedersen quoted in Frellesvig (Frellesvig, 1996, p. 98)):

When, in the most essential, most carnal, most indispensable and original words, the foundation of the language, there are similarities between two languages, and then a sufficient number to allow for the formulation of rules for the changes in letters from one to the other, then there is a basic kinship between these languages.

Compare it to the quotation by a) William Jones b) Edward Sapir. What do they have in common? How do they differ?

2. **This is how Joseph Greenberg describes the classification of languages by Humboldt (1960, p. 181):**

Passing over other writers who discussed the topic in much the same terms as Schlegel, we come to Wilhelm von Humboldt who, in his essay "Ueber die Verschiedenheit der Menschlichen Sprachen" (1836), placed this type of analysis at the very heart of his approach to language. Von Humboldt viewed each language as a distinct self-revelation of the spirit (Geist). Such self-revelations, while each a valid expression in its own right, exhibit lesser or greater degrees of perfection. There are four classes of languages in von Humboldt's scheme. He adds a fourth, incorporating type to the by now traditional threefold classification in order to accommodate certain American Indian languages whose very complex word-patterns include instances in which the object of a verb is incorporated in the same word as the verb root. Von Humboldt is explicit in rejecting any historical evolutionary interpretation in which higher types evolve out of lower types. These are ideal types involving different degrees of the unfolding of form. The isolating

⁸ For more information on morphological classification of languages cf. chapter 10.

languages are "formless," the incorporating languages, through their over-elaboration, betray no true sense of form. As might be expected, only the inflected languages, by their harmonious fusion of root and affix in a true unity, are credited with a true sense of form.

How would you interpret Humboldt's classification of languages?

3. Study the following scheme and comment on it.

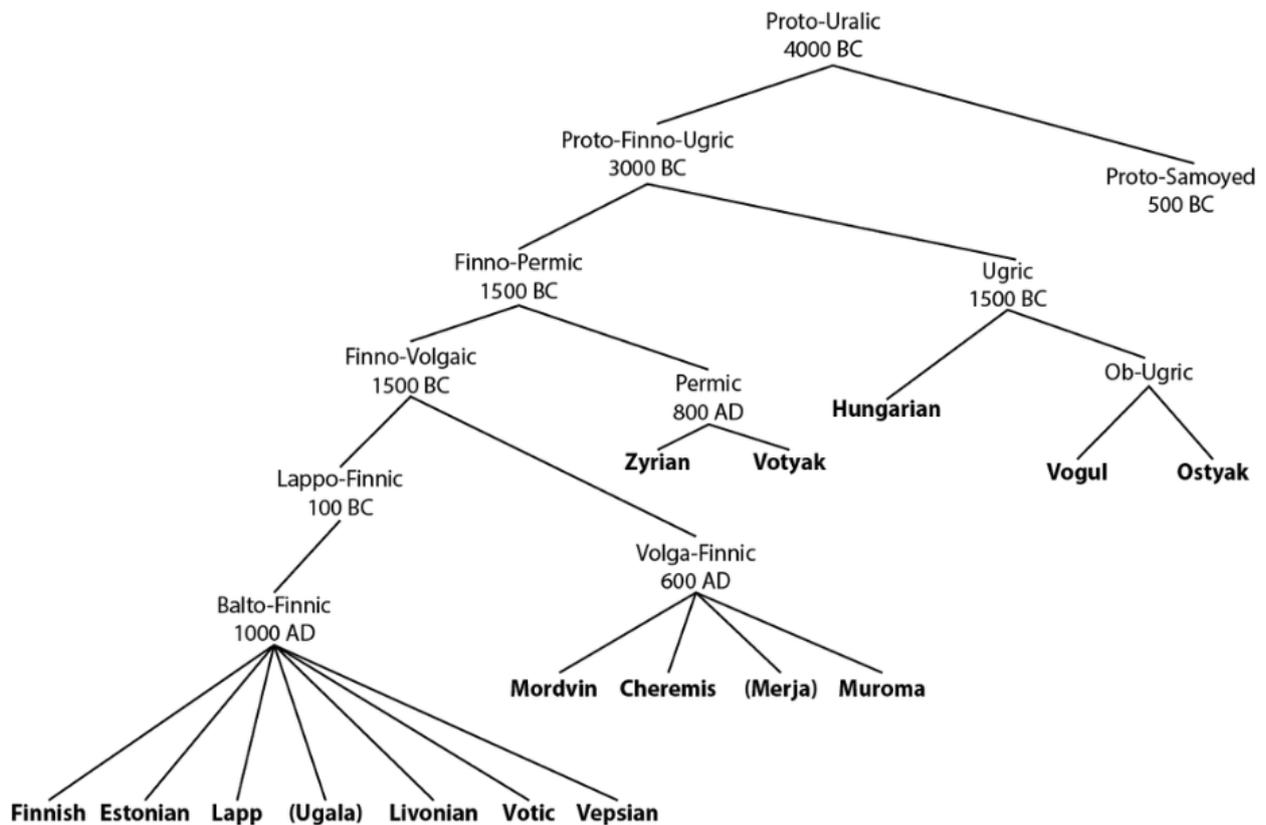


Figure 3 Uralic language family

Source: <https://deepbaltic.com/2016/05/14/neighbours-but-not-relatives-translating-between-estonian-and-latvian/>

Summary

How did language typology start?

What is historical linguistics?

What is comparative linguistics?

What is the Stammbaumtheorie?

Who were the Junggrammatikers?

What is the difference between the holistic and partial approaches to language typology?

Genealogical classification of languages

What is synchrony and diachrony?

What do you remember about August Schleicher?

What are the geographical borders of Europe? What do you know about the history of nations that live in Europe? Where are they from?

Since Saussure two basic approaches to the study of language have been distinguished: the synchronic and the diachronic ones. While synchronic linguistics studies language at a particular point of time, diachronic linguistics focuses on the development of language through the course of time. Any level of the language system can be viewed from these two angles. The diachronic and the synchronic perspectives can also be applied to the field of linguistic typology. In this chapter we will focus on the diachronic perspective.

As discussed in Chapter 2, the development of languages and their comparison was of deep interest among linguists of the 18th century. We mentioned August Schleicher who, being a botanist, compared human language to a living organism. Another metaphor was used to refer to languages that were proved to be of common origin – **language family**. The study of the history of human families is called genealogy and by analogy **genealogical (or genetic) classification** of languages is based on the common history of languages. Genealogy employs family trees to explain how family members are related to one another. Analogical representation can be used to illustrate relationships among languages belonging to one and the same language family. The relationship among ‘sister’ languages is of genetic (or genealogical) nature. The genetic principle underlies the so-called genetic (or genealogical) classification of languages.

The starting point of the genealogical classification are linguistic similarities among languages that follow from their common origin. Based on these similarities languages are grouped into language families. Thus, a language family refers to a group of languages that are related. They descended from a common ancestor – the proto-language of that family. Genetic similarities among languages are explored by the historical-comparative method. This method enables typologists to reconstruct the developmental stages of the individual languages of a language family back to the initial mother language, the so-called **proto-language** (*proto* means ‘early’ in Greek). The protolanguage can be reconstructed from available written documents of the daughter languages, e.g. in the case of the Indo-European language family the reconstruction of the proto-language is based on Latin, Sanskrit, Ancient Greek and others. Of course, to identify the linguistic situation dating back many centuries is

a difficult task and since there are no recordings and documents in the proto-language, its perfect reconstruction is not possible. It can happen that the data are insufficient to establish a family or to group existing families into larger units. According to Blake (2009, p. 246) if languages bear some similarities but the evidence is insufficient, the discussed languages can be described as languages belonging to one **stock**. Obviously, languages in one stock share some properties, they resemble each other. As Blake (ibid) states, if this resemblance is based on 10-20% shared vocabulary (in the case of some stocks it is only 5-10%), these stocks can be said to belong to one **phylum**. We can see another metaphor at play here. The Greek word *phýlon* means a group with common ancestry, tribe. It has been traditionally used in biology as a term that, in terms of generalization, ranks one level above the class. It is a major taxonomic division of a taxonomic kingdom and groups together all classes of organisms that have the same body plan. Analogically, languages of one phylum are characterized by similar language structure, e.g. cognate words. In linguistics, phylum is above a language family, it is a superfamily or macrofamily. So, for example, there is the Afro-Asiatic phylum consisting of Semitic, Cushitic, Omotic and other families (Crass, 2009, p. 378).

Languages of one language family are further divided into smaller groups. Various metaphorical terminologies are used to refer to these smaller groups. Some typologists speak of **branches** and sub-branches, in analogy with trees and their branches. The Germanic group of languages, for example, is a branch of the Indo-European language family and it has three sub-branches: The eastern sub-branch (Gothic), the northern sub-branch (Danish, Icelandic, Norwegian, Swedish), and the western sub-branch (German, Frisian, Dutch, Afrikaans, English) (c.f. (Blake, 2009, p. 251). This approach can be found, for example, also in the Ethnologue database.

In contrast to it, WALS subdivides language families into **genera**. The Latin term *genus* 'birth, race, type' is a taxonomic term originally used in biology – a family of species is divided into genera. In a similar vein, a family of languages is divided into genera. Thus, genus is smaller than a language family; it is its subgroup. This term was introduced into the field of typology by Dryer (Dryer, 1989 p. 267): *“First, the languages are grouped into genetic groups roughly comparable to the subfamilies of Indo-European, like Germanic and Romance. I refer to each of these groups as a genus (following a suggestion by Bill Croft), since they are rather analogous to the taxonomic level of genus in biology.”*

Languages belonging to the same genus have a common ancestor and history – the time depth is about 3500 years.

There are also languages that do not display genetic similarities with any other language. These languages are called **isolates**. It is believed that most of these languages are remnants of families. Blake gives the following examples of language isolates (Blake, 2009, p. 249):

- Ainu (spoken in Japan)

- Burushaski (spoken in northern Pakistan)
- Basque (spoken in the Pyrenees)
- Elamite (an extinct language of southwestern Iran; it has been claimed to be related to the Dravidian languages of southern India)
- Japanese and the Ryukyuan dialects (the latter spoken in the Ryukyu Islands of Japan)
- Ket (spoken in the Yenisei Basin, Siberia)
- Korean
- Nivkh (spoken in eastern Siberia, including Sakhalin Island)
- Sumerian (extinct language of Mesopotamia with records from the 3rd millennium B.C.)
- Yukaghir (spoken in eastern Siberia)

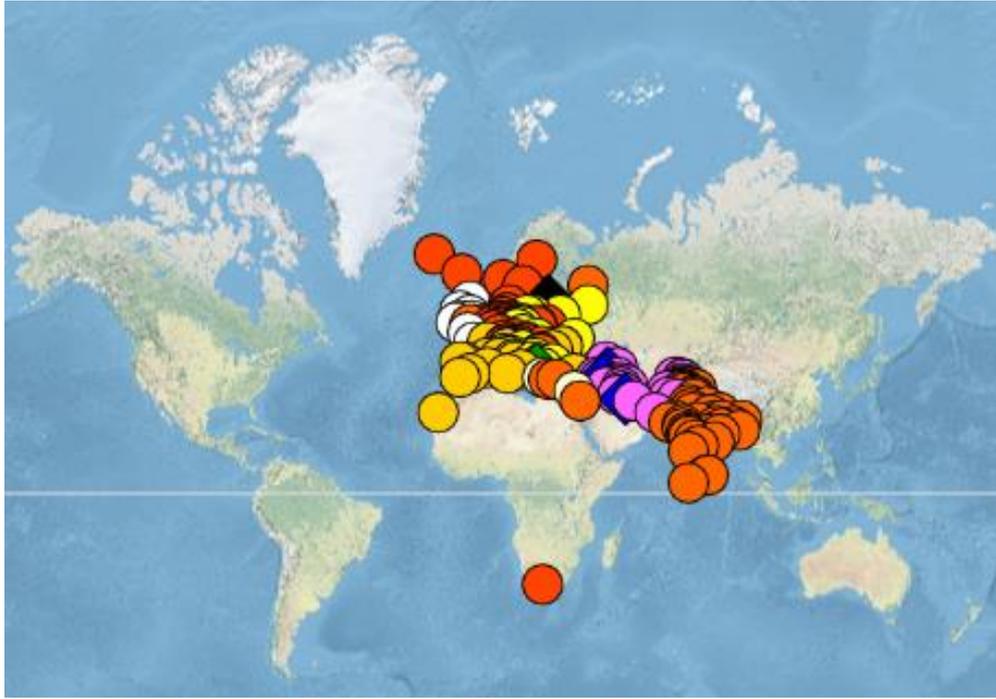
The affiliation of language with a specific language family can be disputed. This might be caused by lack of (written) sources which could be used for comparison of languages. Blake (ibid), for example, also mentions that some linguists include Japanese and Korean in the Altaic family and Yukaghir in the Uralic family.

Generally speaking, each language can be classified from the genealogical point of view. English, for example, is a member of the Indo-European language family and its Germanic genus; Slovak belongs to the same family but to the Slavic genus. According to WALS there are 215 language families. Ethnologue (November 2014) gives the largest language families as follows:

Language family	Number of living languages	Number of speakers
Afro-Asiatic	366	362,281,758
Austronesian	1,221	345,818,471
Indo-European	436	2,916,732,355
Niger-Congo	1,524	430,784,205
Sino-Tibetan	456	1,268,209,279
Trans-New Guinea	475	3,536,267
Totals	4,478	5,327,362,335

Table 11 The largest language families in Ethnologue

The largest language family according to the number of living languages is the Niger-Congo family. It encompasses 1, 524 languages while the Indo-European family consists ‘only’ of 436 languages. On the other hand, the Indo-European language family has the highest number of speakers - 2,916,732,355. The map below illustrates the distribution of language genera in the Indo-European language family. Altogether there are 10 language genera.



Genera

- | | | |
|--|--|---|
| <input type="checkbox"/> + ◆ Albanian (1) | <input type="checkbox"/> + ● Germanic (39) | <input type="checkbox"/> + ● Romance (24) |
| <input type="checkbox"/> + ◆ Armenian (3) | <input type="checkbox"/> + ○ Greek (2) | <input type="checkbox"/> + ● Slavic (17) |
| <input type="checkbox"/> + ◆ Baltic (2) | <input type="checkbox"/> + ● Indic (53) | |
| <input type="checkbox"/> + ○ Celtic (9) | <input type="checkbox"/> + ● Iranian (26) | |

Figure 4 The Indo-European language family (Dryer, et al., 2013)

However, the overall distribution of the individual languages over the world differs from Figure 4. Language territories do not overlap with countries, for example the English language is spoken not only on the British Isles but also in the USA, Canada and Australia. This fact is responsible for the high number of the speakers:

World Language Families *Indo-European*

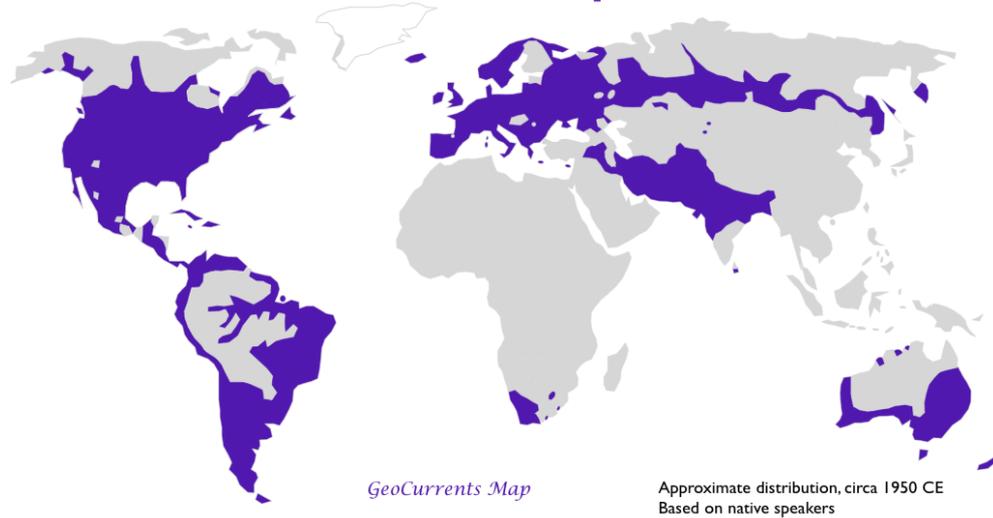


Figure 5 The Indo-European language family in the world

Source: <http://www.geocurrents.info/cultural-geography/linguistic-geography/world-maps-of-language-families/attachment/indo-european-language-family-map>

The map in Figure 5 is interesting for two reasons. First, it illustrates how widespread the Indo-European languages are. This is in contrast to the size of the Indo-European language family (cf. the Table 11 above). Second, this territorial extension is striking if we realize that the Indo-European family originated from one proto-language, originally spoken in one relatively limited area. More precisely, it is supposed that the homeland of the Proto-Indo-European was the region of the Pontic Steppe north of the Black sea and east to the Caspian. This language was spoken more than 6,000 years ago. It (or its dialects) spread over the territories of Europe and Asia.

Contrary to this account, a team of evolutionary biologists from the University of Auckland published a study in 2012, according to which the origin of all modern Indo-European languages is Anatolian, a language spoken on the territory of present-day Turkey (Wade, 2012)⁹. They drew their conclusions on the basis of an experiment that included a list of 103 words that are resistant to linguistic change, e.g. pronouns, parts of the body, and family relations. These words in various modern Indo-European languages were compared to their Indo-European counterparts. The word *mother* is a good illustration of such a word – its cognates (words of the same etymological origin, excluding loanwords) are, for example, the Slovak word *matka*, German *mutter*, Persian *madar* and Latin *mater*. The research team combined the information about cognates with historical and geographical data, and all the data was processed by a computer program.

⁹ cf. http://www.nytimes.com/2012/08/24/science/indo-european-languages-originated-in-anatolia-analysis-suggests.html?_r=0

Comparison of basic vocabulary is typical of comparative linguistics that frequently relies on **lexicostatistics**. It is a quantitative and statistical method focusing on the vocabulary of a language, with special attention paid to historical links with other languages. Basically, this method computes the percentage of common roots in languages. The starting point is the list of basic vocabulary (e.g. Swadesh's list). The higher the percentage the closer the relation among the languages examined. This kind of comparison is not restricted to the lexical level of the language system. In the previous chapter, for example, Grimm's Law was described. It is also known as the First Germanic Sound Shift and as such it formulates the relationship between certain consonants in Germanic languages and their predecessors in Proto-Indo-European. It is a kind of innovation, which separated the Germanic protolanguage from the Indo-European protolanguage. Inflectional morphology can be explored as well, e.g., the genitive ending -s which is common to Greek, Latin and English.

As has already been mentioned, each language can be classified from the genetic point of view. Sometimes, classifications differ since they depend on the scholar's approach or the database used. The same applies to the terminology (cf. branch vs. genus). Since the source of many examples used in this textbook is the WALS website, the following chapters are based on the genetic affiliation of languages as specified in this database.

Tasks

- 1. Have a look at the following years from British history. Identify them with historical events. How did they influence the development of the English language?**

1928
1922
1828
1776
1702
1607
1564
1476
1400
1388
1362
1150
1066
449

- 2. One of the methods for the identification of genetic relationships among languages uses the so-called cognate words. What are those? Can you give an example of cognate words for English, Slovak, German, etc.?**
- 3. The hierarchy of biological classifications includes eight major taxonomical ranks:**

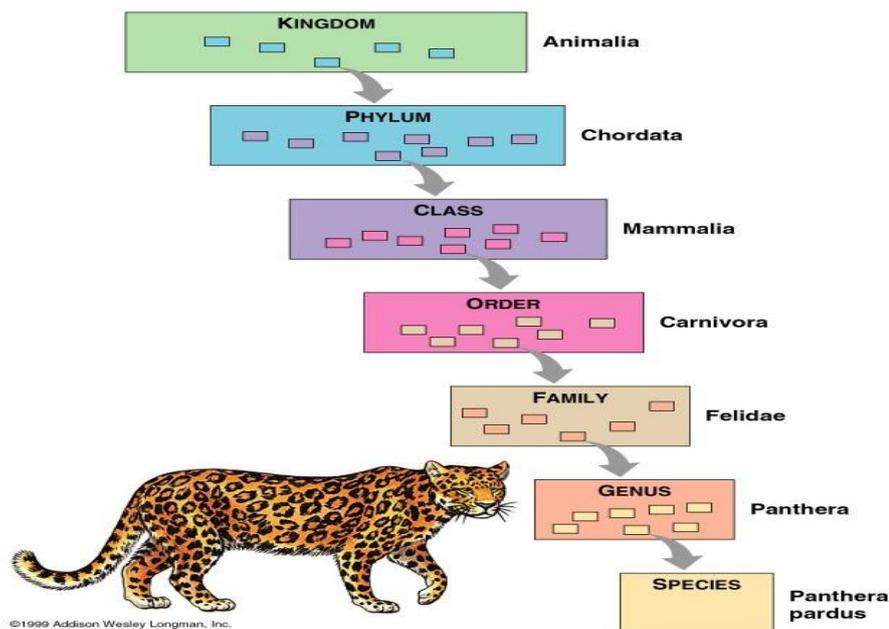


Figure 6 Classification in biology

source: <http://00evevas01873.blogspot.sk/2013/01/dichotomous-key.html>.

What do you prefer as a term labelling a subgroup of languages of a language family? Why?

4. Read the following text and explain the notion of 'genus':

The notion genus is explained in Dryer (1989). It is intended as a level of classification which is comparable across the world, so that a genus in one family is intended to be comparable in time depth to genera in other parts of the world. The choice of term is intended to match the general idea of genus in biological classification, where a genus is a set of species that are clearly closely related to each other (and where words in everyday language often correspond to genera rather than species). In the genealogical classification of languages, a genus is a group of languages whose relatedness is fairly obvious without systematic comparative analysis, and which even the most conservative "splitter" would accept. Genealogical groups deeper than a genus are often less obvious and in the absence of detailed comparative work are often not universally accepted. If there is evidence of time depth of groups, the genus would not have a time depth greater than 3500 or 4000 years. A genus may have a time depth much less than this, but if the time of the split of one group of languages from other languages in the family appears to be greater than 4000 years, then this constitutes a reason to say that this group of languages is a separate genus. The standard subfamilies of Indo-European (e.g. Germanic, Slavic, Celtic, etc.) are fairly clear examples of genera, although Celtic is perhaps a clearer example than Germanic or Slavic, both of which have a time depth considerably less than 3500 years. The decisions as to which groups to treat as genera here are best described as my own educated guesses. In many instances they are based on conversations he has had with specialists. However, in the absence of a tradition within the field of attempting to identify groups of comparable time depth in different parts of the world, they should not be considered more than educated guesses. Specialists who think the choices of genera here are mistaken are encouraged to let me know. (WALS, 2016)

5. Are there any non-Indo-European languages that are spoken in Europe? Are there any language isolates?

6. This is how John Algeo describes the Indo-European culture. Read the text. Can you explain how scientist could arrive at those conclusions?

Indo-European Culture

On the basis of cognate words, we can infer a good deal about Indo-European culture before it spread over many parts of Europe and Asia. That spread started no later than the third or fourth millennium B.C. and perhaps very much earlier. Indo-European culture was considerably advanced. Those who spoke the parent language, which we call Proto-Indo-European, had a complex system of family relationships. They could count. They used gold and perhaps silver also, but copper and iron only later. They drank a honey-based alcoholic beverage whose name has come down to us as mead. Words corresponding to wheel, axle, and yoke make it clear that they used wheeled vehicles. They were small farmers, not nomads, who worked their fields with plows, and they had domesticated animals and fowl. Their religion was polytheistic, including a Sky Father (whose name is preserved in the ancient Vedic hymns of India as Dyaus pitar, in Greek myth as Zeus patēr, among the Romans as Jupiter, and among the Germanic peoples as Tiw, for whom Tuesday is named). The cow and the horse were important to their society, wealth being measured by a count of cattle: the Latin word pecus meant 'cattle' but was the source of the word pecūnia 'wealth,' from which we get pecuniary; and our word fee comes from a related Old English word fēoh, which also meant both 'cattle' and 'wealth.' So we know things about the ancient Indo-European speakers on the basis of forms that were not actually recorded until long after Indo-European had ceased to be a single language.

(Algeo, 2010, p. 50)

Explain the following claim:

“And the absence of a common word for ocean suggests, though it does not in itself prove, that this homeland [of Proto-Indo-European] was inland.” (Algeo, 2010, p. 50)

7. **How was the Indo-European language family discovered? Describe its structure based on WALS and Ethnologue.**
8. **As it was suggested, language families are based on the findings of comparative and historical linguistics. One of the aims has been reconstruction of proto-languages. One of the possibilities is to study similarities and differences in the so-called core vocabulary. What kind of words does the core vocabulary consist of?**
9. **One such word is *father*. Taking into consideration the discussion above, comment on the following scheme. Complete it with the Germanic languages.**

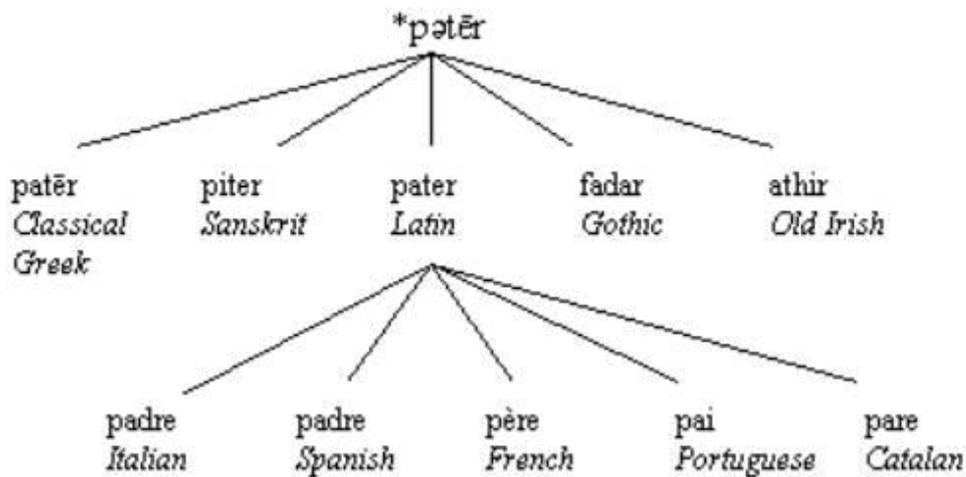


Figure 7 *father* in the Indo-European languages (Tan, 2013)

10. Decode the following information:

English, Germanic, Indo-European, UK.

Try to 'encode' the Slovak language in similar way.

11. Comment on the following words of Edward Sapir in the context of the discussion on language family:

"No two languages are ever sufficiently similar to be considered as representing the same social reality. The words in which different cultures live are distinct worlds, not merely the same world with different labels attached".

12. Go to the WALS database and identify the genealogical classification of the following languages: Hungarian, Maltese, Hausa, Hopi. Go to the Ethnologue and compare the classification. How do they differ?

13. Find examples of isolates in WALS.

14. Have you ever heard about the Nostratic hypothesis? Read the following paragraph and discuss it. Why is it controversial?

As mentioned previously, there can be various degrees of resemblance between language families and the levels of relationship can be quantified lexicostatistically and described in terms of stock and phylum. But besides hypotheses of wider relationships based purely on lexicostatistics, there are hypotheses about possible relationships between families using standard techniques of reconstruction or mixtures of standard methodology and lexicostatistics. The Nostratic hypothesis is one of the boldest and most controversial approaches; largely the work of Aharon Dolgopolsky and Vladimir Illich-Svitych, the hypothesis claims that there is a macrofamily consisting of Indo-European, Semitic, Berber, Kartvelian, Uralic, Altaic, Korean, Japanese, and Dravidian (Dolgopolsky, 1998). (Blake, 2009, p. 249)

Summary

What is the genetic classification of languages?

What a language family?

What is a language phylum?

What is a branch of a language family?

What is a language genus?

What is a language isolate?

What is lexicostatistics?

What is basic vocabulary?

What do you know about the Indo-European language family?

Phonological typology

What is the basic unit of phonology?

Explain the functional view of the phoneme.

Explain the mentalistic view of the phoneme.

Explain the physical view of the phoneme.

Give at least three basic characteristics of the phoneme.

What is the difference between a phoneme and an allophone?

Explain complementary distribution. Give an example.

Explain unidimensional and multidimensional oppositions.

Explain proportional and isolated oppositions.

Explain privative and gradual opposition.

Draw the Cardinal Vowel Scheme.

Explain the criteria for the classification of consonants.

Characterize plosives and fricatives in terms of the place and the manner of articulation.

Explain the difference between consonants and vowels.

What are suprasegmental features?

Explain the following notions: rhythm, stress, tone, mora, pitch.

Language typology involves the comparison of languages. Theoretically, in any comparison there are at least two units – two comparanda and one **tertium comparationis** (Latin = the third [part] of the comparison). Tertium comparationis is the common platform for comparison of two things, the point of comparison. If two dogs are compared and one of them is big while another is small, then being a dog of a specific size is the quality both dogs (big and small) have in common, the so-called tertium comparationis. If we employ Haspelmath's terminology discussed in Chapter 1, the tertium comparationis in language typology is the comparative concept. As Haspelmath points out (2010), comparative concepts are created by typologists for the purpose of comparison and they do not equate to the descriptive formal categories (e.g. past tense, genitive, adjective). Descriptive categories might be similar in languages but they are not the same, they cannot be "captured by equating categories across languages" (Haspelmath, 2010, p. 2). An illustrative example of this situation is the case of the adjective as described by Dixon (cf. Chapter 1). Even though there is a concept of adjective and it can be applied to the description of the majority of languages, its description depends on the specific language.

Comparative concepts can be identified at each level of language system. The following chapters will identify possible points of comparison at four basic language levels: phonology, morphology, syntax and lexicon.

One of the indisputable language universals is that each language has vowels and consonants. Vowels and consonants are combined in various ways to form syllables. Moreover, syllables can differ in rhythm, stress and tone. By implication, phonological typology can focus on basic units (consonants and vowels) and can compare languages from the point of view of e.g. their inventory, or it can deal with possible combinations of consonants and vowels (syllables) and the role of rhythm, stress and tone.

Consonants and vowels

It is a well-known fact that languages may differ in their vowel and consonant inventory. The Hungarian language, for example, has umlaut vowels (ö, ü) that are not present in the Slovak language. On the other hand, the Slovak language has the vowel ä. Let's have a look at the system of vowels in three genetically distinct languages. Sentani is a language of the Sentani language family. WALs classifies it as a language of Papunesia. Alawa is an Australian language and English is an Indo-European language. This is how Cowan describes the system of vowels in the Sentani language (Cowan, 1965 p. 4)

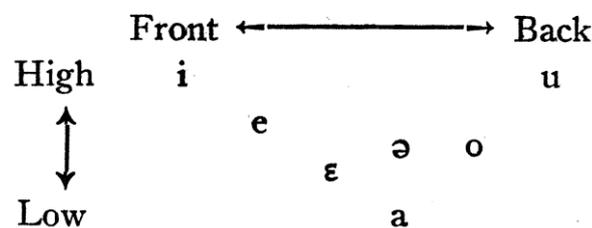


Figure 8 Vowels in Sentani

The vowels of Alawa are as follows (Sharpe, 1972 p. 14):

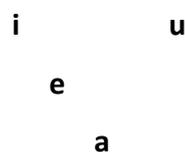


Figure 9 Vowels in Alawa

And as you know, the Cardinal Vowels Scheme of English consists of the following vowels:

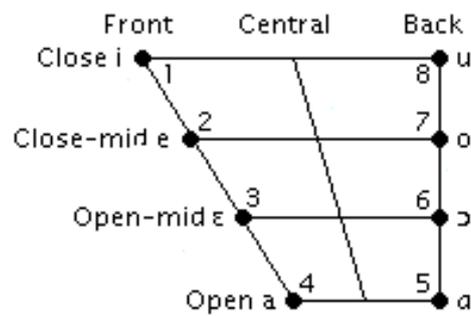


Figure 10 Vowels in English

It is obvious at first sight that these three languages differ in the number of vowels that comprise their vowel inventory. While Alawa has only 4 vowels, English has 8 and Senatni 7. Maddieson (Vowel Quality Inventories, 2013) compared 564 languages from the point of view of vowel quality inventory and divided these languages into 3 groups. Languages with small vowel inventory have 2-4 vowels, an average vowel inventory consists of 5-6 vowels and languages with large vowel inventory make use of 7-14 vowels. Following this typology, Alawa is a language with small inventory, Senatni can be classified as a language with average inventory and the vowel inventory of English is large.

The inventory of consonants can be compared in an analogical way. Generally speaking, the number of consonants is higher than the number of vowels. By implication, the number of types in classification of languages from the point of view of consonant inventory is higher. In total, there are five types (Maddieson, 2013) of them, ranging from 6 to more than 34 consonants:

- Small from 6 to 14 consonants
- Moderately small from 15 to 18 consonants
- Average with 22 ± 3 consonants
- Moderately large from 26 to 33
- Large with 34 or more consonants

Inventories of vowels and consonants in individual languages can be compared and the Consonant-Vowel Ratio can be calculated. Maddieson (Consonant-Vowel Ratio, 2013) claims that even though there is no overall correlation between consonant inventory size and the number of vowel qualities, “by looking at the two aspects jointly a subtler picture of their interaction can be obtained”. English, for example, is a language with a low Consonant-Vowel Ratio.

Syllables

Languages may differ in the way the sounds of a language are combined in a syllable. In the Slovak language, for example, a combination of three and even more consonants is possible, e.g. *zmrzlina*, *krk*, *krb*. These words are usually unpronounceable for an English speaker despite the fact that the individual sounds can be identified in the English language.

The simplest combination is one combining a consonant and a vowel in the C+V/ V+C syllable type. For some languages, the C+V combination is the only one to occur in its system. Maddieson refers to it as to the simple syllable structure (Maddieson, Syllable Structure, 2013). Besides the simple syllable structure Maddieson differentiates between a moderately complex syllable structure and a complex syllable structure. If a language has a moderately complex syllable structure, it means that a single consonant can occur after the vowel and/or two consonants can occur before the vowel. Only the common two-consonant patterns described above are allowed here. Languages “which permit freer combinations of two consonants in the position before a vowel, or which allow three or more consonants in this onset position, and/or two or more consonants in the position after the vowel” (Maddieson, Syllable Structure, 2013) are languages with complex syllable structure. In the Anamuxra language, for example, the following syllable structures are possible (Ingram, 2001 p. 36): V, VV, CV, CVV, VC, VVC, CVC, CVVC, CCV, CCVV, CCVC, CCCVC. This means that Anamuxra is a language with a moderately complex syllable structure.

Prosodic features

The generally recognized prosodic features are rhythm, stress and tone. Languages can be compared, for example, from the point of view of **rhythmical** patterns. In English and German, for instance, there are stressed and unstressed syllables. The unstressed syllables are reduced. In French and Spanish, there are stressed and unstressed syllables too, but the syllables are more equal in duration. In Japanese, rhythm is determined by the metrical weight of sub-syllabic units (Maddieson, Typology of phonological systems, 2013, p. 537). Based on these features, languages can be divided into stress-timed, syllable-timed and mora-timed.

According to Moravcsik (2013, p. 176), in almost half of the world’s languages **stress** is predictable.

(Goedemans , et al., 2013) classifies languages with fixed stress into 7 types:

1. No fixed stress (mostly weight-sensitive stress)
2. Initial: stress is on the first syllable
3. Second: stress is on the second syllable
4. Third: stress is on the third syllable
5. Antepenultimate: stress is on the antepenultimate (third from the right) syllable

6. Penultimate: stress is on the penultimate (second from the right) syllable
7. Ultimate: stress is on the ultimate (last) syllable

Examples of individual language types are as follows (Goedemans & van der Hulst (Fixed Stress Locations, 2013)):

- A fairly large number of languages have initial stress. An example is Cahuilla (Uto-Aztec; California): *'ñāʔa,čeh* 'sit down', *'neñukum* 'female cousins'.
- A few languages have stress on the second syllable. An example is Mapudungun (Araucanian; Chile and Argentina): *ti'panto* 'year', *e'lumu,yu* 'give us'.
- Only one language in our sample, Winnebago (Siouan; Illinois), exhibits stress on the third syllable: *hochi'chinik* 'boy', *waghi'ghi* 'ball' (see also Hayes 1995).
- An example of a language with antepenultimate stress is Paumarí (Arauan; Amazonas, Brazil): *ra'bodiki* 'wide', *oni'manari* 'seagull'.
- Penultimate stress is again fairly common. A language with this pattern is Djingili (West Barkly; Northern Territory, Australia): *bi'angga* 'later', *,ɲuru'ala* 'we all'.
- A language that exhibits ultimate stress is Weri (Trans-New Guinea; Morobe, Papua New Guinea): *u,lua'mit* 'mist', *,aku,nete'pal* 'times'.

Goedemans & van der Hulst (2013) also observe the following correlations between geography and fixed stress locations.

- European systems (if not weight-sensitive) mostly have initial stress. The percentage of weight-sensitive languages is greater in the south.
- Austronesian systems mostly have penultimate stress.
- Arabic dialects are commonly weight-sensitive.
- South American languages are mostly weight-insensitive, with stress oriented towards the right edge of the word.
- North American systems form a rather heterogeneous group.
- Except for the northern part (where we find the non-Pama-Nyungan languages with penultimate stress), Australian systems mainly have initial stress.

From the point of view of **tone**, languages are generally divided into tonal and non-tonal languages.

Tone refers to the use of pitch to express lexical and grammatical meanings. In his study of 527 languages, Maddieson (Maddieson, Tone, 2013) divided the tonal languages into two groups – languages with a simple tonal system and languages with a complex tonal system. The simple tone system means that a language makes use of a two-way basic contrast, usually between the high and the low levels. The contrast set of languages with a complex tonal system is more intricate. According to Maddieson's (Tone, 2013) findings "languages without tones predominate in the western part of

the Eurasian landmass, including South Asia, in the more southerly regions of South America, and in the coastal area of north-western North America. In this last area great genealogical diversity exists among the indigenous languages, but tone is almost entirely absent. In addition, no Australian language has been reported to be tonal". Languages of Africa are usually tonal and they usually have simple tonal systems. Languages with complex tone systems are typical of East and Southeast Asia.

Tasks

1. Go to the Ethnologue database and find as much information as possible on Sentani and Alawa. Complete this information using the WALS database. Describe the procedure.
2. Find the chapter on vowel inventories in WALS. Find the map and comment on the areal distribution of individual vowel inventories. Are there areas with high concentration of small/average/large inventories? Identify at least three languages with small, average and large vowel inventories. Where does Slovak/English belong?
3. Find the chapter on consonant inventories in WALS and identify one language per each type.
4. Read the chapter in WALS on Consonant-Vowel Ratio by Ian Maddieson. How does it classify the languages? How is the ratio calculated? What is typical of the geographical distribution of individual C/VQ ratios?
5. Find in WALS the chapter on syllable structures and comment on the distribution of individual language types. Where does English/Slovak belong?
6. The consonantal system of Anamuxra consists of sixteen underlying phonemes which are built around four points of articulation: bilabial, alveolar, velar and palatal (Ingram, 2001 p. 20):

		Bilabial	Alveolar	Palatal	Velar
Obstruents	Oral	p, β	t, s		k, ɣ
	Prenasal	b	d, z		g
Nasals		m	n		ŋ
Liquid			r		
Glides				y	w

Figure 11 Consonants in Anamuxra

The Sentani consonants are (Cowan, 1965 p. 5):

	plosives	fricatives	nasals	lateral	semivowels
labials	b	f	m		w
gingivals	d		n	l	
prepalatals					j
velar	k				
laryngeal		h			

Figure 12 Consonants in Sentani

Compare both systems to English.

7. This is the description of tone in Chinese:

“The first tone is high level. It is relatively constant in its intensity (loudness) and somewhat longer than in tones 2 and 4, but shorter than tone 3.” (Norman, 1988 p. 147)

How many tones are there in Chinese? What type does it belong to? What about English/Slovak?

8. Charney claims that the structure of a typical Comanche noun or verb root is CVS.

However, following combinations are attested too (Charney, 1994 p. 25):

CVV	paa	water
VCV	ata	uncle
CVCV	papi	older brother
VCVCV	akw Vsi	sneeze (noun)
CVCVCVV	p̄imotōō	cow
VCVCVCV	anikúta	ant
CVCVCVCVV	wakaté?ee	turtle

Table 12 Syllable structure in Comanche

What type of syllable structure is it?

9. Work with the WALS database and answer the following questions:

- What is the minimal inventory of vowels/consonants?
- How many vowels/consonants form the smallest inventory? Find an example.
- What is the most common vowel inventory? Give examples of languages with the most common vowel inventory.

- d) What is the most typical consonant inventory? Give examples of languages with the most typical consonant inventory.
- e) Is the majority of world languages of tone type?

Summary

Name possible tertium comparationis at the sound level.

What are comparative concepts? Do they overlap with descriptive categories?

Describe language typology from the point of view of: consonants, vowels, syllables, stress, tone.

Morphological typology

Characterize morpheme and allomorph. What is the difference between a word and morpheme?

Explain the following notions: cranberry morph, clitic morph, empty morph.

Explain the notion of paradigm.

Explain the notion of conjugation.

Explain the notion of declension.

Explain the difference between root and stem.

Explain three basic criteria for the classification of word-classes.

What is the difference between referential (natural) gender and formal gender? What kind of gender is typical of English?

Explain the difference between tense and time.

Explain the difference between finite and non-finite verb forms. Which forms are non-finite?

Explain the difference between perfective and progressive aspects.

What is the difference between word-formation and morphology?

Explain the notions determinant vs. determinatum.

What is the difference between semasiology and onomasiology?

Explain derivation. Give examples.

What is the difference between blending and compounding?

What is the difference between conversion and derivation by zero morpheme?

What is moneme and what is zero morpheme?

What is the difference between back-formation and noun incorporation?

What is the difference between clipping and acronymization?

The discussion on the typological classification of languages from the morphological perspective is a very complex issue. In chapter 2 that discusses the history of typology, it is mentioned that the Schlegel brothers, August Schleicher and Wilhelm von Humboldt classified languages on the basis of their internal structure. In the majority of cases it was the morphological description of a language that was decisive for its classification. Even though there are many reservations about this classification today, attributes such as analytic, agglutinative, inflective, etc., are still commonly used to describe individual languages. The next factor that influences the discussion on the morphological classification of languages is the position of morphology in the language system. Any discussion of morphological (and any other) typology crucially depends on the delimitation of the scope of the field in question (here, the field of morphology). For these two reasons, this chapter is divided into several subchapters.

First, the morphological classification established by the Schegel brothers is discussed. Second, the possibilities of classifying languages from a morphological perspective are introduced in more detail.

Morphological classification of languages (holistic approach)

As previously mentioned, the underlying goal of various comparative studies, especially in the 19th century, was the reconstruction of the protolanguage. It was believed that comparing words and their morphological structure in various languages could result in reconstruction of the Indo-European protolanguage. At the same time, linguists tried to identify similarities and differences between languages. They compared and contrasted the morphology of words and grouped languages with similar morphological features into one class. The morphology of individual languages served as a tool for the classification of complete language systems (i.e., not only their morphological level). This approach is labelled as the **holistic** approach to language typology. The linguists that established the holistic tradition have already been mentioned. Here comes their summary:

The first to come up with the idea of classification of languages was **Wilhelm Friedrich von Schlegel** (1808 p. 45). He based his typology on grammatical criteria and distinguished between two types of languages: languages that express relations between words by inflection – **flexional** languages and languages that use affixes for the same purpose – **affixal** languages. He did not draw a clear distinction between them, but the difference may be expressed as “the simple combination of morphemes vs. the phonological alternation of morphemes in combination” (Croft, 1991 p. 45). His older brother, August von Schlegel,¹⁰ elaborated on this typology and added the third type – languages that do not have (inflectional) morphology. By implication, the following three types of languages, based on morphological criteria, were distinguished:

- affixal languages (languages with affixes)
- flexional (languages with inflections)
- languages with no structure (languages without inflectional morphology)

Their contemporary, **Wilhelm von Humboldt**, focused not only on the morphological structure but also on the structure of sentences (1988, p. 168): “Far more, however, than in individual words, the intellectual diversity of nations is exhibited in the *constructions of speech*, in the range that they are able to give to *sentences*” He was deeply interested in the idea of a perfect, ideal language. In contrast to his counterparts, his research also covered languages spoken on the American continent. He distinguished three possible forms (Humboldt W. , 1988, p. 216):

¹⁰ Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie grecque-latine-allemande.

- **flectional** (relations between words in a sentence are expressed by affixes)
- **isolating** (lack of inflection)
- **incorporative**¹¹ (sentence is reduced to a noun form, verb and noun form one word)

The only exception is Chinese which, according to Humboldt (1988, p. 216), lacks any grammatical forms – there are no grammatical forms and word order is of great significance. Humboldt also maintains that the presence of one property indicates the presence/absence of some other thus indicating the importance of searching for correlations. Furthermore, he is aware of the fact that one language can demonstrate properties of various language types.

This threefold typology of languages as well as parallels between language development and Darwin's theory of evolution found their reflection in the work of **August Schleicher**. Schleicher refers to Darwin in relation to a genealogical classification of languages and he also proposes the family tree structure (*Stammbaum*, cf. Chapter 2). Schleicher excludes incorporating languages from his typology and works with:

- **isolating** languages (languages without affixes)
- **agglutinating** languages (affixes denote single grammatical categories, are joined together one after another with little phonological alternation)
- **inflectional** languages (languages that use affixes, often fuse grammatical categories into one affix, this fusion may be accompanied by phonological alternations) (Croft, 1991 p. 46).

Schleicher identifies three stages in the development of languages. First, there is language birth. Language is born as an isolating language and it further develops into an agglutinative type. Finally, it reaches the final stage of its development – the inflectional type. The inflectional type occupies the highest position in this hierarchy.

The Schlegel brothers, Humboldt and Schleicher all have one thing in common: they tried to match the whole language with one language type. The basis of this classification included morphological and syntactic properties of a language. A breakthrough in the approach to typological classification is connected with the work of **Edward Sapir**. He points out that although properties of languages can be generalized and covered by one umbrella type, a criterion for this generalization must be chosen. Based on studies of various languages of the world, he suggests a set of criteria for the classification of languages. Sapir distinguishes between *technique* (formal process) and the *degree of synthesis* (how many morphemes are used in one word) and calls his criteria indexes.

Index of synthesis

¹¹ Nowadays we prefer the label 'polysynthetic'.

The question asked in the case of the index of synthesis is how much syntactic information is obtained in the average word; how many morphemes are there in a word? Three language types are distinguished respectively¹²:

- **Analytic**: almost every word consists of one morpheme, the number of grammatical affixes is small; languages may or may not have derivational affixes. Word order and function words convey syntactic information:

Vietnamese: Khi tôi đến nhà bạn tôi, chúng tôi bắt đầu làm bài.
 when I come house friend, I begin do lesson
 When I came to my friend's house, we began to do lessons.

- **Synthetic**: a word consists of a small number of morphemes, words consist of more morphemes, they are combinations of the root/stem and grammatical morphemes:

Slovak: *Eva kúpila knihu.*
 Eva buy-3SG:PAST book-ACC.SG
 'Eva bought a book.'

- **Polysynthetic**: words tend to consist of several morphemes, a verb and an object are incorporated into one word:

Yup'ik Inuit: *tuntussuqatarniksaitengqiggtuq*
 tuntu -ssur -qatar -ni -ksaite -ngqiggte -uq
 reindeer -hunt -FUT -say -NEG -again -3SG:IND
 'He had not yet said again that he was going to hunt reindeer.'

Index of fusion

The question asked is: How do morphemes usually build words? There are three possibilities:

- **Isolating** languages: The number of affixes of any kind is limited. All isolating languages are usually analytic.
- **Agglutinative** languages build words by adding several derivational or inflectional affixes glued one after the other:

Chukchi: *t*E -meyN*E -levt*E -p*E*ft -*Erk*En*
 1 SG -great -head -ache -IMPF
I have a fierce headache.

- **Fusional** languages build words by adding affixes, but usually not more than one or two in a single word:

Slovak: *robot-ník*
 work- AG
 'worker'

¹² The source of examples in the following section is the Glossary of linguistic terms (<http://www.01.sil.org/linguistics/GlossaryOfLinguisticTerms/contents.htm>) that collected the examples from various sources.

- **Symbolic** languages build words by modifying the root rather than adding affixes, e.g.:

English: tooth-PL
'teeth'

run-PAST
'ran'

Based on these indexes, a classification should be viewed as a continuum or a scale. If we focus on the number of morphemes in a word (index of synthesis) at the one end of the scale there are analytic languages with a one-to-one relation between a word and morpheme. On the other hand, there are polysynthetic languages that combine morphemes into long linear sequences.

If the focus is on the segmentation of words into morphemes (index of fusion) a similar situation can be observed. At the one end of the scale there are isolating languages, at the other end fusional languages. By implication, languages cannot be identified entirely with one type.

Sapir's typology was further developed by **Joseph Greenberg**, who took a quantitative approach to the problem – which makes his theory unique. Greenberg identifies various parameters that are calculated as “a ratio of two units, each defined with sufficient rigor and by calculation of a numerical index based on the relative frequency of these two units over stretches of text”. (Greenberg, 1960 p. 185) Greenberg identifies the index of synthesis, index of agglutination, compounding index, derivational index, gross inflectional index, prefixal index, suffixial index, isolational index, pure inflectional index, concordial index. He selected 100 word long passages for 8 languages, including two ancient Indo-European languages, Anglo-Saxon and Sanskrit, and analysed and compared them in respect to individual indices. English and Persian served him to illustrate the changes in time. The Annamite language was picked as an example of an isolating language, Eskimo exemplified a polysynthetic language, Swahili and Yakut exemplified agglutinative languages. All in all, Greenberg's method defines different types of languages on the basis of the above-mentioned indices.

As it has been suggested, analytic languages overlap with isolating languages. In a similar vein, synthetic languages can be viewed from the point of view of index of fusion as either agglutinating or fusional. The term *fusional* is often replaced by **inflectional**. This typology into isolating/analytic, synthetic (divided into agglutinating and inflectional), polysynthetic and introflective (symbolic) languages is preferred nowadays. Even though typology is based on morphemes, their number and function, word order in sentences, plus additional phonological make-up, it is still used to refer to a whole language. In the following section each language type is characterized and exemplified.

LANGUAGE TYPE: analytic/isolating

Languages: Mandarin Chinese, Vietnamese, Cantonese, Cambodian, Thai, Ewe, Gungbe, English

Features:

- words consist of single morphemes – mostly of a root
- words are monosyllabic and monomorphemic
- few derivational/ inflectional morphemes
- tendency to form words by combining free morphemes into compounds
- sentences consist of a series of free morphemes
- relatively fixed word order
- extensive use of function words
- less rigid grammatical rules
- the grammatical meanings of person and number in conjugation of verbs is not expressed by suffixal morphemes, instead personal pronouns are used, which thus function as grammatical morphemes
- declension is not realized by special case morphemes but by prepositional morphemes separated from lexical verb form (father, of the father, to the father, father)
- lack of agreement e.g. in declension of possessive syntagmas

Example:

Gungbe (Huttar, et al., 2012 p. 17):

Ùn ná wlé àjòtó [dě fin hùn cè] fè
1sg fut catch thief rel steal drum 1SG.POSS certainly
'I will certainly catch the thief [i.e., whoever that could be] who stole my drum.'

LANGUAGE TYPE: synthetic - inflectional

Languages: Greek, Latin, Sanskrit, Russian, Slovak, Polish

Features:

- words consist of stem and affixes
- affixes often mark several grammatical categories simultaneously (cumulation)
- one inflectional morpheme can carry more than one set of morphological functions (syncretism)
- primary means of building new words is by adding affixes
- rich in morphological synonymy/homonymy/polysemy

Example: Slovak: the singular declension paradigm of *chlap* 'man'

Nominative	chlap-0
Genitive	chlap-a
Dative	chlap-ovi
Accusative	chlap-a
Locative	chlap-ovi
Instrumental	chlap-om

Table 13 Slovak declension paradigm

LANGUAGE TYPE: synthetic - agglutinating

Languages: Finnish, Hungarian, Estonian, Swahili, Turkish

Features:

- words consist of a stem and one or more clearly identifiable affixes
- addition of a large number of affixes one after another
- primary means of building new words is by adding affixes
- every morpheme has just one meaning
- one to one relationship – one form – one meaning
- Vowel harmony is typical.
- auxiliaries are rarely used
- lack of grammatical synonymy and homonymy
- simple grammar, regular
- possession is not expressed by special pronouns but by possessive grammatical
- modal meanings are very often expressed synthetically

Example: Marad-hat-ok a nagy ház-ban.

stay -can -I ART big house-DAT.Sg.

'I can stay in a big house'

LANGUAGE TYPE: polysynthetic

Languages: Chukchi, Ainu, Mohawk, Cheyenne

Features:

- words consist of long strings of stems and affixes, which may translate as an entire English sentence
- extensive use of inflection, derivation and compounding
- express grammatical categories, derivational and lexical meanings in one word.
- noun incorporation – especially noun by verb which results in very long verbs

Example:

Yupik (de Reuse, 2006 p. 746):

Atan aangelghimeng qikmilguuq.
 ata- -n aange- lghii-
 father ABS.2S.S be.big INTRANS.PARTL
 -meng qikmigh- -lgu- -uq
 ABL.S dog have.N INDIC.3SG
 'Your father has a big dog'.

LANGUAGE TYPE: introflective

Languages: Hebrew, Arabic, and other Afro-Asiatic languages

Features:

- internal changes replaced by inflections
- only partial agreement and gender
- the word consists of a consonantal skeleton – usually 3 consonants, so called radicals that expresses a general basic meaning, so called primary meaning. Inserted vowel gives the word a secondary meaning
- synonymy and homonymy more frequent than in agglutinating languages but less frequent than in inflectional languages
- besides the inflectional type, the most anomalous type of language

Example: German Vater: Väter; Mutter: Mütter,

English: sing-sang-sung; tooth-teeth; goose-geese

The term 'introflection' refers to the nonconcatenative morphological processes, typical of e.g. Semitic languages. The root consists of a set of consonants – the meaning of this root is usually general, e.g. *g-d-r* in Hebrew means 'enclose' (Glinert, cited in Velupillai, 2012 p. 98). Vowels are inserted between the consonants to express grammatical meanings:

Past	a-a	(CaCaC)	gadar	'enclosed'
Present	o-e	(CoCeC)	goder	'encloses'
Future	Yi-O-o	(yiCCoC)	yigdor	'will enclose'
imperative	O-o	(CCoC)	gdor	'enclose!'
Infinitive	li- O-o	(liCCoC)	ligdor	'to enclose'

Table 14 Introflection in Hebrew

The list of those who developed morphological typology would not be complete without **Vladimir Skalička**. As a representative of the *Prague School of Linguistics* he represents the holistic approach to linguistic typology: on the basis of selected grammatical features he matches a language with a particular type. Basically, Skalička follows the tradition of the 19th century linguistics. He distinguishes the following types: agglutinative, inflectional, isolating (analytic), polysynthetic and

introflective. However, the description of these types slightly differs from the traditional one, especially in the case of polysynthetic languages.

From holistic to partial approach

A different perspective of morphological typology is to focus on the internal structure of languages and to compare, analyse, and contrast various morphological processes in individual languages. The aim of this comparison is not to classify a language system as a whole; instead, this approach compares the morphological level of various languages. While the former approach classifies languages into holistic types, the second deals with partial morphological categories and processes.

The word 'morphology' was introduced to the field of linguistics by Goethe in analogy to natural sciences where morphology was a branch of biology dealing with the study of the form and structure of organisms and their specific structural features. In linguistic morphology two types of processes can be distinguished – processes that change the form of a word and processes that change both the form and the meaning. There is no clear borderline between these two types and therefore the relation between them should be understood as a continuum. The nature of the first type of process is inflectional, and morphology dealing with inflectional processes is therefore labelled as inflectional morphology. The second type of morphological processes is derivational, and that's why the label derivational morphology is used in this case. Another label used is word formation¹³. Thus, traditionally the morphological level can be divided into inflectional morphology and derivational morphology.

There are various options for a tertium comparationis at the morphological level. Languages can be compared on the basis of categories, as for example, case, gender or number, or on the basis of word classes (and respective categories). So, for example, the number of cases can be compared. Iggesen (Iggesen, 2013) compares the number of cases in 261 languages and identified the following types:

- no morphological case-marking
- 2 case categories
- 3 case categories
- 4 case categories
- 5 case categories
- 6-7 case categories
- 8-9 case categories
- 10 or more case categories
- exclusively borderline morphological case-marking

¹³ These labels are used synonymously in this textbook, even though there are authors who do not include compounding in word-formation, e.g. Anderson (1992) (1993), Lees (1960), Aronoff (1976).

Similarly, the number of genders can be compared. Corbett (Corbett, 2013) proposes five types of languages according to this criterion: languages without gender, languages which distinguish 2, 3, 4 genders and languages with 5 and more genders.

These examples of language typologies are available in the WALS database. However, there is only one study in WALS that deals with derivational morphology – the chapter on reduplication by C. Rubino. In contrast to numerous typological studies in the field of inflectional morphology, the number of publications on typology in derivational morphology / word-formation is very low. One of the rare cases in which word formation is discussed from the point of view of classification of languages is Štekauer, Valera and Körtvélyessy (2012). They compare 55 languages representing 33 language families. 1500 examples illustrate similarities and differences among languages that are evaluated in terms of the presence vs. absence of a word-formation process or feature. The analysis enables the authors to formulate some generalizations, e.g.:

- Every living language has means to coin new linguistic signs.
- Cross-linguistically, the hierarchy of word-formation processes in terms of their frequency of occurrence is as follows: suffixation, compounding, prefixation, conversion, back-formation, circumfixation, infixation, etc.
- Suffixation is cross-linguistically the most widespread word-formation process
- There is a strong tendency for languages with derivational suffixation to make use of recursive suffixation, i.e., more than one derivational suffix is permitted.

Štekauer (2012) suggests a three-fold word-formation typology of languages:

1. The first criterion is the **nature of expansion** of the word-formation processes. Based on this, the word-formation processes can be prevalingly **expansive** or **non-expansive** and three types of languages can be distinguished: **concatenation** type (majority of languages), **non-concatenation** type (e.g. East-Dangla), **mixed** type – both concatenative and non-concatenative processes play roughly the same role in the word-formation system of a particular language (Konni, Hausa, Cirecire).
2. The second criterion classifies the languages according to the prevailing **word-formation strategy**. If the language is of concatenation type, there are 4 options: 1) the language combines free morphemes. This includes compounding, incorporation, full reduplication. 2) free morphemes are combined with bound elements (affixation processes, partial reduplication – West Greenlandic, Lakota; 3) both strategies are combined. The majority of languages belong to this type. 4) Words are formed without free morphemes. This covers root and patterns, blending, neo-classical compounding, e.g. in Hebrew. If the language is of non-concatenative, it is either of conversion type, back-formation type or stretch/pitch

modification type. It should be, however, mentioned that no language of our sample belonged to this prevalingly non-concatenation type.

3. The last identified criterion was the **richness** of productive word-formation processes. Languages are of two types in this respect: 1) **derivationally rich** – it means that they make productive use not only of the major but also of the minor word-formation processes (English, Slovak, Karao, Hebrew); 2) languages are **derivationally poor** – it means that their word-formation capacity is restricted to a minimum or they make use of a limited number of word-formation processes – e.g. Tatar, Lakhota, West Greenlandic. Štekauer distinguishes 3 classification criteria: the prevailing type of word-formation processes; the prevailing type of word formation strategy and the richness of productive word formation processes.

Generally speaking, this is the first classification of languages from the perspective of word-formation. Moreover, it is based not only on presence vs. absence of a word-formation process in the word-formation system of a language but it reflects the **structural richness** of word-formation processes.

Tasks

1. **Morphological typology goes back to Friedrich Schlegel. He identified 2 language types. This is how Joseph Greenberg translated the description of one type: (Greenberg, 1960 p. 180)**
“heap of atoms which every wind of chance scatters or sweeps together”
The following sentence describes one of the language types by Friedrichs’s brother, August:
“One might say that all their words are roots, but sterile roots which produce neither plants nor trees” (ibid)
Can you explain the metaphors used by Schlegel brothers? What language types are described by them?
2. **This is how two centuries later William Croft interpreted language types by F. Schlegel. Read the descriptions and match them with language types. Which part of the sentence refers to the ‘heap of atoms’ above?**
“the simple combination of morphemes vs. the phonological alternation of morphemes in combination.” (Croft, 1991 p. 45)
3. **According to August von Schlegel there are 3 types of languages: affixal, flecional and languages with no structure. Explain these notions. The following characteristics can help you:**
 - languages with affixes

- languages with inflections
- languages without inflectional morphology

4. According to Humboldt there are 3 types of languages: flecional, isolating, incorporative.

Explain these notions. The following characteristics can help you:

- relations between words in a sentence are expressed by affixes
- lack of inflection
- sentence is reduced to a noun form; verb and noun form one word

5. Schleicher excluded the incorporating type of language from his typology, and worked with isolating, agglutinating and inflectional languages. Describe them with the help of the following characteristics:

- languages without affixes
- affixes denote single grammatical categories, are joined together one after another with little phonological alternation
- languages that use affixes, often fuse grammatical categories into one affix, this fusion may be accompanied by phonological alternations (Croft, 1991 p. 46)

6. Read the paragraphs taken from Sapir. What should language typology be based on? Vocabulary or language structures?

So far, in dealing with linguistic form, we have been concerned only with single words and with the relations of words in sentences. We have not envisaged whole languages as conforming to this or that general type. Incidentally we have observed that one language runs to tight-knit synthesis where another contents itself with a more analytic, piece-meal handling of its elements, or that in one language syntactic relations appear pure which in another are combined with certain other notions that have something concrete about them, however abstract they may be felt to be in practice. In this way we may have obtained some inkling of what is meant when we speak of the general form of a language. For it must be obvious to anyone who has thought about the question at all or who has felt something of the spirit of a foreign language that there is such a thing as a basic plan, a certain cut, to each language. This type or plan or structural "genius" of the language is something much more fundamental, much more pervasive, than any single feature of it that we can mention, nor can we gain an adequate idea of its nature by a mere recital of the sundry facts that make up the grammar of the language. When we pass from Latin to Russian, we feel that it is approximately the same horizon that bounds our view, even though the near, familiar landmarks have changed. When we come to English, we seem to notice that the hills have dipped down a little, yet we recognize the general lay of the land. And when we have arrived at Chinese, it is an utterly different sky that is looking down upon us. We can translate these metaphors and say that all languages differ from one another but that certain ones differ far more than

others. This is tantamount to saying that it is possible to group them into morphological types. (1921, p. 127)

I have in mind valuations of form as such. Whether or not a language has a large and useful vocabulary is another matter. The actual size of a vocabulary at a given time is not a thing of real interest to the linguist, as all languages have the resources at their disposal for the creation of new words, should need for them arise. (Sapir, 1921, p. 131)

7. Based on the following examples identify the language type. What type of language is English?

- I will come at noon.
- The nicest weekend I have ever had was in late September two years ago.
- anti-dis-establish-ment-arian-ism

8. This is how Greenberg calculated the degree of synthesis (Greenberg, 1960 p. 185):

The first of these parameters is the degree of synthesis or gross complexity of the word. Since Sapir's time the minimum meaningful sequence of phonemes in a language has come to be called the morpheme in American linguistics. For example, the English singing contains two morphemes but forms one word. The ratio M/W where M equals morpheme and W equals word, is a measure of this synthesis and may be called the synthetic index. Its theoretical lower limit is 1.00, since every word must contain at least one meaningful unit. There is no theoretical upper limit, but in practice values over 3.00 are infrequent. Analytic languages will give low results on this index, synthetic higher, and polysynthetic the highest of all.

These are the values for the index of synthesis. What do they suggest?

Language	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Index								
Synthesis	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72

Table 15 Index of synthesis calculated by Greenberg

9. What type of language is it? Choose from among the following types: analytic, inflectional, polysynthetic, agglutinative and introflective.

Language type 1:

- ✓ words consist of a stem and one or more clearly identifiable affixes
- ✓ addition of a large number of affixes one after another
- ✓ primary means of building new words is by adding affixes
- ✓ every morpheme has just one meaning
- ✓ one to one relationship – one form – one meaning
- ✓ vowel harmony is typical

- ✓ auxiliaries are rarely used
- ✓ lack of grammatical synonymy and homonymy
- ✓ simple grammar, regular
- ✓ possession is not expressed by special pronouns but by possessive grammatical morphemes
- ✓ modal meanings are very often expressed synthetically

Language type 2:

- words consist of single morphemes – mostly of a root
- words are monosyllabic and monomorphemic
- few derivational/ inflectional morphemes
- tendency to form words by combining free morphemes into compounds
- sentences consist of a series of free morphemes
- relatively fixed word order
- extensive use of function words
- less rigid grammatical rules
- the grammatical meanings of person and number in conjugation of verbs is not expressed by suffixal morphemes, instead personal pronouns are used, which thus function as grammatical morphemes
- declension is not realized by special case morphemes but by prepositional morphemes separated from lexical verb form (father, of the father, to the father, father)
- lack of agreement e.g. in declension of possessive syntagmas

Language type 3:

- words consist of long strings of stems and affixes, which may translate as an entire English sentence
- extensive use of inflection, derivation and compounding
- express grammatical categories, derivational and lexical meanings in one word
- noun incorporation – especially noun by verb which results in very long verbs

Language type 4

- ✂ internal changes replace inflections
- ✂ only partial agreement and gender
- ✂ the word consists of a consonantal skeleton – usually 3 consonants, so called radicals that express a general basic meaning, so called primary meaning. Inserted vowel gives the word a secondary meaning

- ✦ synonymy and homonymy more frequent than in agglutinating languages but less frequent than in inflectional languages
- ✦ besides the inflectional type, the most anomalous type of language

Language type 5:

- ❖ words consist of stem and affixes
- ❖ Affixes often mark several grammatical categories simultaneously (cumulation)
- ❖ one inflectional morpheme can carry more than one set of morphological functions (syncretism)
- ❖ primary means of building new words is by adding affixes
- ❖ rich in morphological synonymy/homonymy/polysemy

10. Find examples of individual language types according to the number of grammatical cases. Study the respective map in WALS and comment on the geographical distributions.

11. The chart below illustrates the use of word-formation processes in the sample of 55 languages. Comment on it.

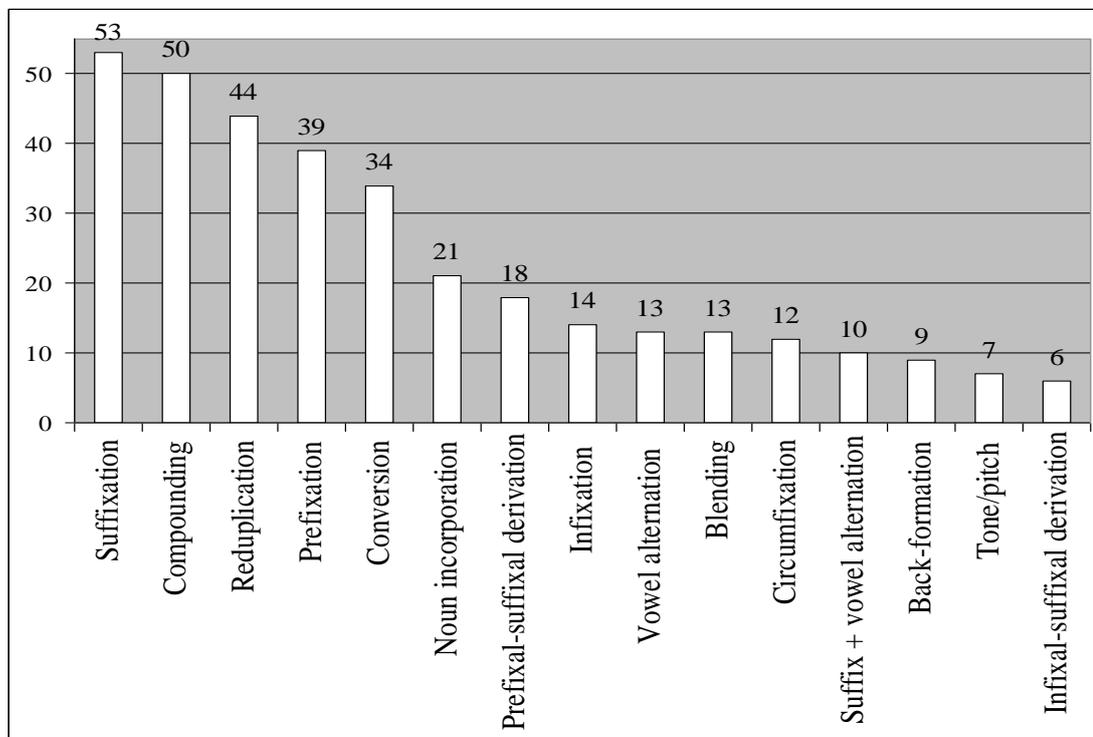


Figure 13 Cross-linguistic use of word-formation processes in the research sample (absolute values with respect to fifty-five languages)

Summary

Name two possible approaches to morphological typology.

What is holistic typology? What language types do you know? Describe them.

What is partial typology?

Give an example of typology in the field of inflectional typology.

What do you know about typology in the field of word-formation?

Syntactic typology

How can syntax be approached?
Describe the Chomskyan approach.
Describe the Bloomfieldian approach.
Describe FSP. – What is this?
What is a sentence?
What is a syntagm?
What are simple sentences?
What are complex clauses?
What are sentence elements? How can they be ordered?
How are questions formed? Compare Slovak and English.
What is negation?

Syntax is a branch of linguistics that studies arrangements of words in linear sequences. On the other hand, typology classifies languages into language types on the basis of shared features. Thus, the combination of syntax and language typology results in a typology of word arrangements that occur in a syntagm.

There are various subjects to be studied – particular constructions, e.g. *phrasal constructions* such as verb phrases (to eat an apple) or noun phrases (Peter's car), *clausal constructions* such as negations and comparisons, or *sentential constructions*, e.g. questions, reported speech, word order, etc. Thus, syntactic typology focuses on phrase structures, clause structures and sentence structures. It studies e.g. the order of individual constituents, relationships between individual constituents and correlations between them. Syntactic typology classifies human languages into types on the basis of shared syntactic properties. In the following subchapters three big areas of syntactic typology will be discussed: order of sentence elements, relationships between syntactic elements and clause types.

Order of sentence elements

The typology of word order studies the arrangement of words in phrases, clauses and sentences. The topic covers not only the basic word order but also, for example, the respective order of adjective and noun, numeral and noun, demonstrative and noun etc. Let us start with the word order of basic sentence elements. There are three basic sentence elements: Subject, Object and Verb. Three major constituents can be combined into six logical permutations: SOV, SVO, VSO, VOS, OSV, OVS. Based on

an analysis of simple, indicative sentences, word order in individual languages can be matched with one of the six possibilities:

SVO	English	<i>The dog</i>	<i>chased</i>	<i>the cat.</i>
		The dog	chase. PAST	the cat
		S	V	O
SOV	Basque	Jon	sendagileari	joan zaio.
		Jon	doctor.DAT	go AUX
		S	O	V
		Jon	went	to the doctor.
VSO	Welsh	<i>Gwelodd</i>	<i>Rhiannon</i>	<i>ddraig.</i>
		see.PAST.3S	Rhiannon	dragon
		V	S	O
		Rhiannon	saw	a dragon.
VOS	Tukang Besi	<i>No-‘ita-‘e</i>	<i>na kene-no</i>	<i>te ana.</i>
		3R-see-3OBJ	NOM friend-3POSS	CORE child.
		V	O	S
		The child	saw	its friend.
OVS	Urarina	<i>obana</i>	<i>itɔafwa-a</i>	<i>katɕa</i>
		collared peccary	shoot-3PS/A	man
		O	V	S
		The man	shot	the collared peccary.
OSV	Tobati	<i>Syaw mahai</i>	<i>nehu</i>	<i>mo-ikor-i</i>
		paddle big	1SG	CAUS-snap-3SG
		O	S	V
		I	broke	the big paddle

Table 16 Examples of word order¹⁴

The word-order typology reflects the dominant type of word-order. By implication, other types of word-order can be expected in respective languages. Besides these six language types there are also languages that lack dominant word order, e.g. Dutch, Hupa or Luiseño.

A particular word order may suggest other properties of a language. For example, English is a typical SVO language. This kind of language prefers prepositions to postpositions (*on the desk vs the desk on*). Udihe, on the other hand, is an example of a SOV language. As the example below illustrates, it prefers postpositions:

Mo: *due-digi-ni eugi-e mi.*

tree top-ABL-3SG come.down-PAST-1SG

‘I came down from the tree.’ (Nikolaeva & Tolskaya, 2001, p. 401)

¹⁴ Source of the examples for individual languages in the table: Basque: (Hualde & Ortiz de Urbina, 2003, p. 449); Welsh: (Borsley, Tallerman, & Willis, 2007, p. 33); Tukang Besi: (Donohue, 1999, p. 65); Urarina: (Olawsky, 2006, p. 168); Tobati (Donohue, Mark. 2002. Tobati. In Lynch, John and Ross, Malcolm and Crowley, Terry (eds.), The Oceanic Languages, 186-203. Richmond: Curzon. p.196).

Similarly, English adjectives are placed before a noun (*a nice man* vs. *man nice*). In Udihe the order is the opposite:

Bi ise:-mi tege-we ño:mi bi:-we-ni.
 Me see.PAST-1SG gown-ACC blue be.PRP-ACC-3SG.
 'I saw a blue gown.' (Nikolaeva & Tolskaya, 2001, p. 178)

Another interesting correlation can be found between the dominant word order and a preference for suffixes/prefixes in inflectional morphology. While suffixation tends to correlate with the SOV word order and postpositions, the SVO order correlates with little affixation and prepositions.

Relationship between syntactic elements

This label covers various topics, e.g. the linkage of syntactic elements, the relationship between them, negation and questions.

How are the syntactic elements linked?

What seems to be a very simple question in one's native language starts to be rather complicated as soon as it is compared to typologically different languages. In English, for example, two nouns in one syntactic phrase (nominal conjunction) are usually joined by 'and'. When it comes to verbs (verbal conjunction), they are joined in the same way:

Nominal conjunction: *Peter and Paul like reading.*

Verbal conjunction: *Paul likes reading and listening to music.*

The same can be observed, for example, in Navajo, Mohawk, Khoekhoe, Ket, Russian, Tagalog and Spanish. However, for example in Somali, the situation is different (the example is taken from WALS, chapter 64):

Nominal conjunction	rooti	iyo	khudrat	
	bread	and	fruit	
	'bread and fruit'			
Verbal conjunction	Wuu	cunnay	oo	cabbay.
	FOC.3M.SG	eat	and	drink
	'He ate and drank'.			

Table 17 Nominal and verbal conjunction in Somali (Berchem, 1991, pp. 324-27)

The same pattern can also be observed in Slave, Teribe, Cantonese and Chocho.

How is the relationship between syntactic elements expressed?

The next frequently discussed question pertaining to the relationship between syntactic elements is the way of expressing the relationship between sentence elements. This question is intriguing especially in languages without dominant word order and in languages in which the word order is not fixed. One of the typological criteria concerns the opposition between the so-called pro-drop and non-drop languages. In pro-drop (or pronoun-dropping) languages, unstressed pronouns are normally dropped and the person and number of the subject are encoded in the inflection of the verb (e.g. in Slovak and Polish). In non-pro-drop (no pronoun dropping) languages, pronouns are used to express grammatical categories, as for example in French and English, e.g.:

Slovak: *Čítam knihu.*

English: *I read a book.*

Other topics under discussion are the agreement of sentence elements, expression of possession, types of predication etc.

Negation

Another widely discussed topic is negation, especially the position and the expression of the negator. There are two basic questions that correlate: How is negation expressed (what kind of morpheme is used to express negation)? and Where is the negator placed in the sentence? The following possibilities are identified by Mathew Dryer (WALS): negative affix, negative particle, negative auxiliary verb, negative word, unclear if verb or particle, variation between negative word and affix, double negation (for examples see Task 7).

Questions

This topic can be exemplified by the typology of polar questions. Polar questions are also called yes/no questions. It means that the answer to these questions starts either with YES or NO. Construction of this kind of questions varies in individual languages. German or English, for example have interrogative word order, it means that a different word order is used to express the question, e.g. *Are you reading?* Some languages simply change the intonation of the sentence (Armenian) or use special question particle, e.g. in French: *Est-ce que vous aimez le livre? Do you like the book?*

Clause types

In WALS there are 4 chapters dealing with different clause types. Each chapter starts with the following explanation: *This chapter examines (investigates) the form of the verb in XYZ clauses.* Obviously, the verb form decides the typology of a language from the perspective of specific clause type. Basically,

clauses are parts of complex sentences, which means that at least two verb forms can be discussed. The author of the chapters in question, Sonia Cristofaro, applies Stassen's (1985) distinction to classify verb forms in purpose clauses as either balanced or deranked. A balanced verb form is one that can occur in an independent declarative clause. Cristofaro exemplifies it by a Mandarin Chinese sentence (the part of the sentence in the brackets is a purpose clause):

Nǐ guì-xialai [qiú Zhāng-san]

You kneel.down beg Zhang-san

'You knelt down in order to beg Zhang-san.' (Cristofaro, 2013, p. from (Li and Thompson 1973: 98))

The linked verbs in the sentence above show the same morphological form. At the same time, each of them could occur in an independent declarative clause. This is not true of the English translation of the sentence. The verb in the main clause is in past form and as such it could occur in an independent declarative clause. The second verb, *beg*, however, is in infinitive form and as such it is structurally different from the verb in the main clause. By implication, this verb form is deranked. Deranked verb forms may display, among other things, special tense, aspect, mood or person markers not used in independent declarative clauses.

Thus, the typology of clauses can be twofold – balanced and deranked. In addition, there is the third type of languages in which both deranked and balanced verb forms exist.

Tasks

1. Read the text below. What does syntactic typology study?

"Syntactic typology is concerned with discovering cross-linguistic patterns in the formation of particular constructions, whether those constructions be phrasal (e.g. noun phrase possessives, equivalent to John's house), clausal (e.g. basic content questions, equivalent to What did Marvin eat?) or sentential (e.g. the reporting of speech, equivalent to Phoebe says that she enjoys swimming). There are two basic descriptive goals when comparing such structures as they arise in a diverse set of languages. First, the comparison is aimed at identifying the full range of devices that languages employ to create the structures. Second, it determines the relative frequency with which a specific device, or a specific constellation of devices, is used in languages. When these goals have been successfully met, then, the syntactic typologist can answer the question of just how different languages can be in the way they form a construction (be it possessives, content questions, or anything else) and what set of linguistic devices is more (or less) likely to be used for this purpose.

Beyond these core descriptive goals, syntactic typology established, whenever possible, correlations between the mechanisms that a language utilizes in the formation of a construction and other linguistic properties of the language, or between mechanisms and the geography or genetic classification of the language. " (Whaley, 2013 p. 465)

2. What can be studied in syntax? Below you can find four different areas of syntactic studies given in WALS. In the box there are various topics that can be discussed within individual areas. Match the areas with topics.

Passive Constructions; ‘When’ Clauses; Number of possessive nouns; Order of Subject, Verb and Object; Possessive classification, Polar Questions; Genitive, adjectives and relative clauses; Reciprocal Constructions; Order of Adjective/numeral/genitive and Noun; Adjectives without nouns; Order of Degree word and adjective; Reason Clauses; Noun phrase conjunction; Purpose Clauses; Order of Genitive and Noun; nominal and verbal conjunction; Comparative constructions

Nominal Syntax:

Word order:

Simple Clauses:

Complex Clauses:

3. Study the map on basic word order in WALs. Can you observe any correlations between the word order and geography? Can you suggest any correlations between word order and genetic classification? How does word order correlate with the position of case affixes? Identify the type of word order in the following languages (the source of the examples is WALs, chapter 81 by Mathew S. Dryer). Where are the languages spoken? What is their genealogical classification?

Nias

i-rino vakhe ina-gu

i-rino vakhe ina-gu
 3SG.REALIS-cook ABS.rice mother-1SG.POSS
 ‘My mother cooked rice.’

Mandarin

Zh āngs ān shōudǎo-le yi- fēng xìn.

Zh āngs ān shōudǎo-le yi- fēng xìn.
 Zhangsan receive-PERF one-clf letter
 ‘Zhangsan received a letter.’

Irish

Léann [na sagairt] [na leabhair].
 Léann [na sagairt] [na leabhair].
 read.PRES the.pl priest. pl the.pl book.pl
 The priests are reading the books.

Hixkaryana

toto y-ahos i-ye kamara
 toto y-ahos i-ye kamara
 man 3:3-grab-DISTANT.PST jaguar
 ‘The jaguar grabbed the man.’

4. Interpret the following statement. Go to WALs and create a map of the suggested correlation.

As it is stated in Song (Song, 2001 p. 119)“...languages with VO and/or PrN orders have both prefixes and/or suffixes in their morphology, whereas there is a strong tendency for languages with OV and/or NPo orders to possess only suffixes. This can be most clearly observed when languages with exclusive suffixing or prefixing are considered: if a language has exclusive

prefixing, it almost always has PrN and VO orders, not NPo and OV orders, or if a language has NPo and /or OV order, it has exclusive suffixing, not exclusive prefixing”.

PrN – preposition noun order; NPo – Noun-postposition order

5. Can you explain the correlation between SVO word order, prepositions and minimum affixation?

6. Have a look at the following examples. What do they suggest about the junction of noun phrases and verb phrases in the given languages? (the source of the examples is WALS, chapter 64 by M. Haspelmath).

(1) **Hungarian**

- a. *Anna és Péter*
Anna and Péter
- b. *Péter olvas és tanul.*
Péter reads and studies
- c. *Péter olvas és Anna tanul.*
Péter reads and Anna studies

(2) **Dagbani**

- a. *doo ηɔ mini m Ba*
man this and my Father
'this man and my father'
- b. *O biε ka kɔyisi.*
he be.bad and be.thin
'He is bad and thin.'
- c. *Gbuɣima ηubiri nimdi ka jansi Diri kɔdu.*
lion.PL chew.IMPf meat and monkey.PL eat.IMPf banana
'Lions eat meat and monkeys eat bananas.'

7. Classify the types of negations in declarative sentences in the following languages. The options are (the task is based on WALS, chapter 112 on Negative morphemes by Mathew S. Dryer):

Negative affix, Negative particle, Negative auxiliary verb, Negative word, Unclear if verb or particle, Variation between negative word and affix, Double negation

(1) **Grebo**

- a. *ne¹ du¹-da² bla⁴*
1SG.SUBJ pound-PST.BEFORE.YEST rice
'I pounded rice before yesterday.'
- b. *ne¹ yi²¹-da² bla⁴ du¹ mɔle*
1SG.SUBJ NEG-PST.BEFORE.YEST rice pound Monday.
'I did not pound rice on Monday.'

(2) **Musgu**

à sādà cécébè pàý
3SG.M know jackal NEG
'He didn't see the jackal.'

(3) Finnish

e-n syö-nyt omena-a
NEG-1SG eat-PTCP apple-PART
'I didn't eat an apple.'

(4) Kolyma Yukaghir

met numö-ge el-jaqa-te-je
1SG house-LOC NEG-achieve-FUT-INTR.1SG
'I will not reach the house.'

(5) Rama

a. nkiikna-lut uut aa kain-i
man-PL dory NEG make-TNS
'The men don't make a dory.'

b. *i-sik-taama*

3-arrive-NEG
'He did not arrive.'

(6) French

Je ne vois pas la lune.
1SG NEG see.1SG NEG the moon
'I do not see the moon.'

8. Use WALS online and:

- identify the section on Order of sentence elements. Read the relevant chapter and study the map. Summarize the research.
- Find evidence for the claim about the correlation between suffixation, SOV and postpositions.
- Go to the section on negation, find the map and comment on it.
- Identify those topics in WALS (not mentioned in the chapter above) that contribute to the discussion on the relationship between sentence elements.
- Identify four chapters in WALS on clause types mentioned in the text above. What clause type (balanced, deranked, both) prevails?

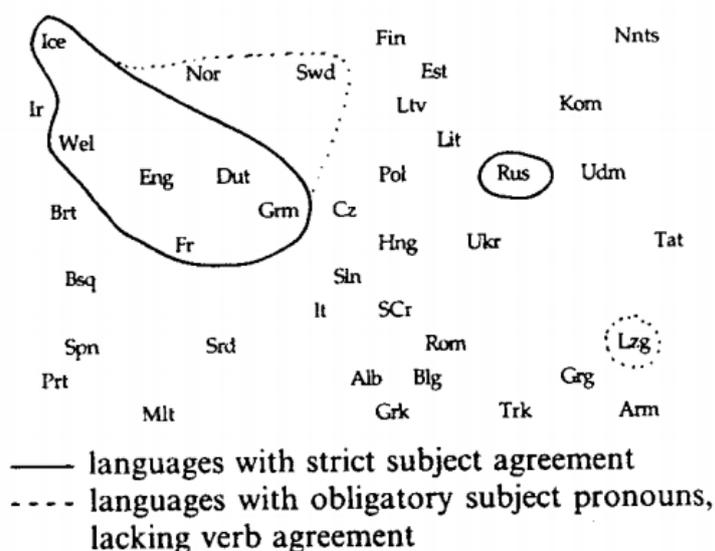
9. Read the following paragraph and comment on the map:

2.11. Subject person affixes as strict agreement markers

*The majority of the world's languages have bound person markers on the verb that cross-refer to the verb's subject (or agent). When these subject affixes co-occur with overt subject NPs (full NPs or independent subject pronouns), they are called agreement markers. However, in most languages they can occur on their own and need not co-occur with overt subject NPs. For example, in the Bulgarian phrase *vie rabotite* 'you (pl.) work', we see the subject suffix *-ite* (2nd person plural) co-occurring with the independent subject pronoun *vie* 'you (pl.)', showing that *-ite**

is an agreement marker. But in Bulgarian it is equally possible and probably more common to say just *rabotite* 'you (pl.) work', i.e. the subject suffix can have a referential function on its own. In German, by contrast, this is not possible: 'you work' is *ihr arbeit-et*. Since the agreement suffix *-et* does not have such an independent referential function, the subject pronoun *ihr* cannot be omitted. Languages like German are often called "non-pro-drop languages", and languages like Bulgarian are called "pro-drop languages"; better terms would be "strict-agreement languages" vs. "referential-agreement languages".

It has sometimes been thought that strict agreement, as exhibited by German, English, and French, is the norm and that referential agreement is somehow special. But in fact, referential agreement is far more widespread in the world's languages, and strict subject agreement is characteristic of a few European languages, some of which happen to be well-known. In her world-wide sample of 272 languages, Siewierska (1999) finds only two strict-agreement languages, Dutch (an SAE language) and Vanimó (a Papuan language of New Guinea). Siewierska further notes that outside of Europe, she is aware of only two additional strict-agreement languages that are not in her sample (Anejóm and Labu, two Oceanic languages). Gilligan (1987) reached a similar conclusion on the basis of a sample of 100 languages. The distribution of strict subject agreement markers in some European languages is shown in Map 107:11.



Map 107.11: Obligatory subject pronouns

Figure 14 Obligatory subject pronoun

The map shows two non-contiguous areas in which subject agreement suffixes cannot have a referential function: Germanic and Gallo-Romance languages with Welsh on the one hand, and Russian on the other. Perhaps only the western European area should be thought of as being relevant for SAE; in Russian, past-tense verbs do not have subject person affixes, so Russian is not a very good example of a strict-agreement language. In the eastern Nordic languages (Norwegian, Swedish, Danish), the subject pronouns are obligatory as they are in English, German or Icelandic, but the languages have lost agreement distinctions on the verb entirely (cf' Swedish *jag biter/du biter/han biter* 'I/you/he bite(s)', Icelandic *ég bit/pú bitur/hann bitur*). These languages are thus "non-pro-drop" in a sense, but they are not strict-agreement languages. English is approaching this type, as the only remnant of subject agreement is the 3rd person singular present-tense suffix *-s*. (There are also some languages of this type) ... (Haspelmath, *The European linguistic area: Standard Average European*, 2001)

Summary

What are the areas of study in syntactic typology?

What is the typology of languages from the perspective of word order? How does word order correlate with suffixation/prefixation? How does word order correlate with the position of the adjective?

How can sentence elements be joined?

How can languages express negation?

Discuss possible ways of question formation.

How can clauses be classified?

Lexical typology

Explain the difference between lexical unit vs. lexeme, word-form and citation form.

Explain the difference between synonymy and homonymy.

Explain the difference between homonymy and polysemy.

Explain three types of homonyms, give examples.

What is the difference between converseness and conversion?

Explain the difference between homonymy and conversion.

Explain hyperonym, hyponym, and co-hyponym.

What is a semantic field?

What is the lexicon?

Explain the difference between bound and free forms.

What is the difference between metaphor and metonymy?

Explain componential analysis, give an example.

What is complementarity?

Explain the principles of IC-analysis.

Explain suppletion.

The main focus of lexical typology study can be identified by simple analysis of the term itself. *Lexical* suggests a relation to the lexicon, where lexicon is understood as either all the words and lexical units used in a language or that a particular person knows. By implication, *lexical* could be explained as relating to the lexicon; relating to individual units stored there. If this explanation is matched with the notion of typology, lexical typology can be understood as a systematic cross-linguistic study of language lexicons; the classification of languages into types on the basis of their lexicons. Even though the definition sounds simple, it is obvious that it covers demanding comparison of lexicons of numerous languages, either contrastive or comparative. The objective is to identify and generalize similarities and differences on the basis of selected criteria. The most frequently quoted definition of lexical typology comes from Lehrer (Lehrer, 1992, p. 249) who understands lexical typology as the identification of “characteristic ways in which language [...] packages semantic material into words”. By ‘semantic material’ we understand various meanings that, from the linguistic point of view, are usually expressed by means of linguistic signs. The most frequently studied and analysed domains in lexical typology are kinship terms, body parts, colour terms and motion verbs.

One of the leading personalities in lexical typology, Maria Koptjevskaja-Tamm, suggests three different perspectives that can be taken to lexical typology: **onomasiological**, **semasiological** and the perspective of the **lexicon-grammar borderline** (Koptjevskaja-Tamm, 2012, p. 374). Let’s have a

look at each of these perspectives and illustrate by examples taken from a project conducted by Koptjevskaja-Tamm, labelled as *Linguistics of temperature* (cf. http://temperature.ling.su.se/index.php/Main_Page). The main question of the project is how various languages express central temperature terms (hot, cold, to freeze, etc.). For example, languages like English, Russian, and Swedish etc. have elaborate systems with six or more terms (e.g., hot, warm, lukewarm, chilly, cool, cold, freezing, ...). On the other hand, there are languages with three or only two terms, e.g. the Mwotlap language has a term meaning hot/warm and a term expressing the idea of coldness/coolness: *sɛw* 'hot/warm' and *mɔmjij* 'cold/cool' (Koptjevskaja-Tamm, 2012, p. 375).

The onomasiological perspective

The onomasiological approach suggests that the direction of research is from the meaning or concept to the form. The starting point is extra-linguistic reality and its categorization in individual languages. We ask how individual languages categorize extra-linguistic reality, how the world around us is categorized into domains. If, e.g. TEMPERATURE is such a domain, then the categorization of temperature differs in languages like English and Mwotlap. Koptjevskaja-Tamm (2012, p. 374) asks the following questions:

- What temperature concepts are encoded as words across languages, what distinctions are made in the systems of temperature terms and what factors underlie them?
- Are temperature terms and temperature term systems completely free to vary across languages, or are there limits to this?
- Are there universal temperature concepts, e.g. 'hot', 'cold', 'temperature'?
- How can the meanings of temperature terms be described?

The semasiological perspective

The semasiological approach offers a reverse perspective – we start with the form to end up with the meaning. Thus, the starting point is the word, its form, and by its analysis we study its various meanings. The focus is, for example, on different meanings expressed by one lexeme. This approach can be exemplified by two adjectives expressing temperature in Swedish and Russian (cf. (Koptjevskaja-Tamm, *Approaching lexical typology*, 2008)). The Swedish *ljummen* means 'lukewarm'; in Russian it is *teplyj* (теплый – the standard English translation is 'warm'). The Swedish *ljummen* refers to neutral temperatures, not warm but not cold. Similarly, the meaning of Russian *teplyj* is restricted, with the greater portion covered by *ljum(men)*. As Koptjevskaja-Tamm (2008, p. 11) points out, “[f]rom the denotational point of view, the two adjectives are, thus, fairly similar to each other, with the denotational range of *teplyj* slightly exceeding that of *ljummen*”:

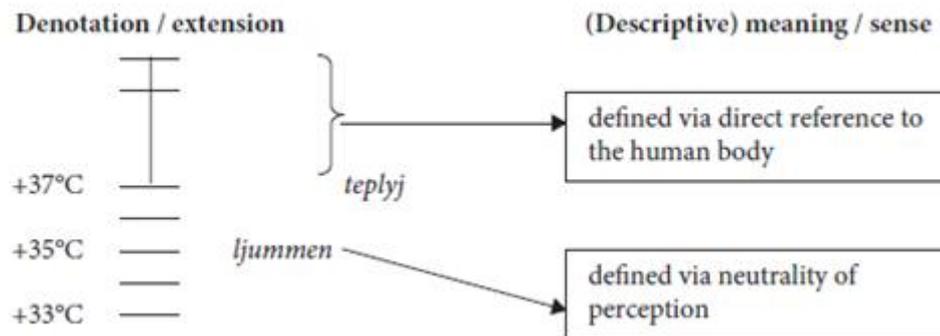


Figure 15 Extension/denotation vs. (descriptive) meaning/sense for the temperature adjectives *ljummen* (Swedish) and *teplyj* (Russian)

However, a comparison of the descriptive meanings of these two adjectives shows big differences. The Russian *teplyj* has a clear orientation towards ‘warm’. It has a comparative form *bole teplyj*, meaning warmer than *teplyj*. On the other hand, such clear orientation cannot be observed with the Swedish adjective. It also has a comparative form but its meaning depends on the context:

Hans öl är ljummare än min.
 (‘His beer is more lukewarm than mine’)
 vs.
Mitt te har svalnat, hans är ännu ljummare
 (‘My tea has cooled down, his is even more lukewarm’).

While in the first sentence the temperature is higher than the standard value *ljummen*, in the second sentence it is lower. Furthermore, in the Russian language, *teplyj* motivates many collocations and metaphors: *teplye slova, čuvstva, otnošeniija* ‘warm (positive, friendly) words, feelings, relations’. These observations lead to the conclusion, that the Russian *teplyj* “designates temperatures that correspond to or are not significantly higher than the temperature of the human body/skin or that maintain the temperature of the human body without too much effort on the part of the human being, and therefore cause an agreeable sensation of comfort and cosiness.” (Koptjevskaja-Tamm, 2008, p. 12). On the other hand, the Swedish *ljumen* lacks the reference to the body and evokes neutrality in perception.

In her ‘temperature’ project, Koptjevskaja-Tamm asks the following semasiologically founded questions (Koptjevskaja-Tamm, *New directions in lexical typology*, 2012, p. 374):

- What are the possible semantic extensions of the temperature meanings to other domains and how can these be related to their concrete meanings?
- Where do temperature terms come from?
- How can the meanings of the temperature terms change within the temperature domain itself?

- What general metaphorical and metonymical models underlie the semantic evolution of the expressions related to the temperature domain?

Lexicon-grammar borderline perspective

The lexicon-grammar borderline perspective studies the interaction and correlation between lexicon and grammar. For example, temperature terms can belong to different word classes, e.g. *hot – heat; cold – freeze*. This relation leads to various syntactic structures. The English *cold* is translated to German as *kalt*. However, the meaning of the English sentence *He is cold* cannot be automatically translated as *Er is kalt*. In German, the structure of the sentence is different: *Ihm ist kalt*.

3 different foci of lexical typology

Based on these three different perspectives, Koptjevskaja-Tamm identifies three different foci of lexical typology (cf. (Koptjevskaja-Tamm, 2008, p. 3); (Koptjevskaja-Tamm, 2012)):

- 1. What is a possible word, or what can be meant by a word?** This question focuses, among others, on possible versus impossible words in different languages, different criteria for identifying words and interaction among them, universal vs. language specific restrictions on possible, impossible.
- 2. What meanings can be expressed by a single word in different languages?** This question covers, for example, the issues of lexicalizations and lexicalization patterns, universal vs. language specific lexicalizations, categorization within, or carving up of lexical fields/semantic domains by lexical items, the architecture of lexical/semantic domains (e.g. basic words vs. derived words).
- 3. What different meanings can be expressed by one and the same lexeme, by lexemes within one and the same synchronic word family (words linked by derivational relations or by lexemes historically derived from each other)?** This question focuses on cross-linguistic patterns in the lexicon-grammar interaction. Attention is paid, for example, to various aspects of motivation, e.g. semantic motivation (polysemy, semantic associations/semantic shifts) and morphological motivation (derivational patterns, including compounding).

Furthermore, these questions can be studied from two different angles labelled **local** and **general** lexical typology (Koptjevskaja-Tamm, 2008, p. 6). Local typology is restricted to a particular lexical field (e.g. TEMPERATURE, BODY, MOTION VERBS), a derivational process, or a particular lexical pattern. The goal of the general typology is to reveal “patterns in the structuring of the lexicon that are supposed to have bearing on many essential properties of the language” (ibid). (e.g. the issues of basic vs. non-

basic vocabulary, or suggestions as to how characterize, compare and measure the lexical-typological profiles of different languages.)

It should also be pointed out that besides the above-mentioned dimensions, lexical typology has also a synchronic vs. diachronic dimension that is stronger than in other typologies. The lexicon of a language constantly changes – new words are coined, old words either disappear or their meaning is modified, shifted. There are also historically oriented lexical typology studies that focus their attention on semantic change, grammaticalization and lexicalization.

Natural Semantic Metalanguage (NSM)

A different approach to analysis at the level of lexicon is offered by the Natural Semantic Metalanguage (NSM) theory. It was originated by Anna Wierzbicka and further developed by Cliff Goddard. The authors claim that it is “the most well-developed, comprehensive and practical approach to cross-linguistic and cross-cultural semantics on the contemporary scene.” It has already been applied to over 30 languages from many parts of the world. (Goddard, 2015). Its basic notions are **explication**, **semantic primes** and **semantic molecules**. **Explications** are paraphrases composed in the simplest possible terms. NSM avoids technical terms, logical symbols, abbreviations etc. In contrary, it makes use of expressions from ordinary natural language, e.g. I, YOU, SOMEONE, DO, HAPPEN, THINK, KNOW, GOOD, BIG, BECAUSE). As Goddard (2011, p. 65) puts it, an explication (reductive paraphrase) should faithfully portray the full meaning of the expression being analysed. This kind of explication can also be applied to comparison of languages. These reductive paraphrases can be described as an attempt to reformulate speaker’s words while the reformulation is based on simplification. At the same time, this reformulation consists of the so-called **semantic primes** (a language consisting of semantic primes creates a kind of mini-language or metalanguage). Semantic primes are also called semantic atoms (cf. Wierzbicka (2014, p. 33). They are semantically basic expressions that cannot be further defined. Altogether there are 63 semantic primes (cf. (Goddard, Griffith University, 2015). It is supposed that semantic primes are universal words, i.e. they have a lexical equivalent in each language. This is how Goddard and Wierzbicka illustrate them:

I~ME, YOU, SOMEONE, SOMETHING, PEOPLE, BODY	Substantives
KIND, PARTS	Relational substantives
THIS, THE SAME, OTHER~ELSE	Determiners
ONE, TWO, MUCH~MANY, LITTLE~FEW, SOME, ALL	Quantifiers
GOOD, BAD	Evaluators
BIG, SMALL	Descriptors
THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEAR	Mental predicates
SAY, WORDS, TRUE	Speech
DO, HAPPEN, MOVE	Actions, events, movement
BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), BE (SOMEONE')S	Location, existence, specification, possession
LIVE, DIE	Life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	Time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH	Space
NOT, MAYBE, CAN, BECAUSE, IF, VERY, MORE, LIKE~AS	Logical concepts

Primes exist as the meanings of lexical units (not at the level of lexemes). Exponents of primes may be words, bound morphemes, or phrasemes. They can be formally, i.e., morphologically, complex. They can have combinatorial variants or allollexes (indicated with ~). Each prime has well-specified syntactic (combinatorial) properties.

Table 18 Semantic primes according to Goddard and Wierzbicka (Goddard, Griffith University, 2015)

Thus, semantic primes are understood as universal words. In the NSM approach there are two types of universal words – besides semantic primes (atoms) there are also **semantic molecules**. Since molecules are more complex than atoms, semantic molecules are more complex than semantic atoms, and they can be defined. The table above suggests that SEE and HEAR are semantic primes (in the tradition of NSM, semantic primes are capitalised). They cannot be further defined. However, ‘colour’ and ‘sound’, for example, are further defined in terms of SEE and HEAR. In a similar vein, cat is an animal [m] (in the tradition of NSM, molecules are signalled by [m]) and sparrow is a ‘bird [m] of one kind’. Semantic molecules can be both language specific and universal (Goddard, Griffith University, 2015). The NSM analysis is illustrated here with *sad* and *unhappy* as given by Goddard (2010, p. 466):

X felt sad:

someone X felt something bad

someone can feel something like this when this someone thinks like this:

“I know that something bad happened

I don’t want things like this to happen

I can’t think like this: I will do something because of it now

I know that I can’t do anything”

X felt unhappy:

someone X felt something bad

someone can feel something like this when this someone thinks like this for some time:

“some very bad things happened to me
 I wanted things like this not to happen to me
 I can’t not think about it”
 this someone felt something like this
 because this someone thought like this

On the basis of these explications, Goddard (2010, p. 466) compares *sad* and *unhappy* as follows:

1) Being *unhappy* requires the experiencer to have certain real thoughts (while one can say *I feel sad, I don’t know why*, it would be a little odd to say *I fell unhappy, I don’t know why*. 2) *Unhappy* conveys a stronger negative evaluation, as implied by the fact that it is less readily combinable with minimizing qualifiers like *a little* or *slightly*. 3) *Unhappy* has a more personal character: one can be saddened by bad things that have happened to other people, but if one is *unhappy*, it is because of bad things that have happened to one personally. 4) *Unhappy* does not suggest a resigned state of mind but rather focuses on some thwarted desires. The attitude is not exact active, because one doesn’t necessarily want anything to happen, but it is not passive either. 5). *Unhappy* suggests a state extended in time. All these differences are modelled in the differences between the two explications.

A similar example can be found in Bromhead’s study of ethnographical categories in English and Pitjantjatjara/Yankunytjatjara (2011): On the basis of use of the word *mountain* taken from Collins Wordbanks (e.g. There is a big mountain behind it that reflects the colours wonderfully. . . ; Erwin looked back up at the mountaintop. . . ; The sun in his eyes was slipping down behind the towering mountains.; . . .we may need to climb mountains . . . ; . . .the wispy band of cloud kissing the distant mountain peaks . . . ; . . .snowboarding down steep mountain slopes.) she gives the following explication of the word *mountain*:

[A] a mountain

- a. a place of one kind
- b. places of this kind are very big places
- c. a place of this kind is above the places on all sides of this place
- d. the top[m] of place of this kind is not like all the other parts of these places
- e. when people see a place of this kind,
 they can think that the other parts of this place are on all sides of the top [m], below the top [m]
- f. the top[m] of a place of this kind is very far above the places on all sides of this place
- g. because of this, when people are far from this place they can see this place
- h. people can think about places of this kind like this:
 “there are not many people in places of this kind”

In South Australia there is Mt Woodroffe. English speakers call it *mountain*, while Pitjanthatjara speakers call it *puli*. *Puli* can have twofold explications:

[B] *puli*₁

- a. something of one kind
- b. things of this kind are very hard [m]
- c. people can do things of many kinds with things of this kind (e.g. kill snakes, pile up)

[C] *puli*₂

- a. a place of one kind
- b. places of this kind are big places
- c. a place of this kind is above the places on all sides of this place
- d. the top[m] of a place of this kind is far above the places on all sides of this place
- e. when people see a place of this kind they can think like this:
“there is a lot of rock[m] (*puli*₁[m]) in this place”

The second explication [C] refers to places like the famous landmark Urulu. Thus, both Urulu and Mt Woodroffe are *puli* in Pitjantjatjara while in English Urulu is a *rock* but never a *mountain*.

Tasks

1. One of the studies on lexical typology is the *hand vs. arm* typology. The Slovak *ruka* corresponds with two English words - *hand* and *arm*. Compare Slovak and English from the onomasiological and from the semasiological perspectives.
2. In a similar vein, discuss the so-called *aqua motion* verbs in English and Russian. Study the table below:

	Passive motion (‘float’)	Self-propelled motion of animate figure (‘swim’)	Motion of vessels and people aboard (‘sail’)
English	<i>float</i>	<i>swim</i>	<i>Sail</i>
Russian	<i>plyt’/plavat’</i>		

Table 19 Aquamotion verbs in English and Russian (Koptjevskaja-Tamm, *Approaching lexical typology*, 2008, p. 8)

3. Translate the following sentences into Slovak. How would you translate the English aquamotion verbs?

I wasn't sure if the canoe would float.

She could swim across the lake.

She always wanted to sail around the world.

4. Go to WALS and read the chapter on Basic Colour categories. Study the corresponding map. Comment on this research and answer the following questions: What is the geographical distribution of the number of colours in individual colour systems? What is the most frequent type of language?

This is how **Kay** (2004, p. 242) describes his approach to colour words:

All the analyst needs to do is to take the phenomenally basic colours: black, white, red, yellow, green and blue as primitives and to define other colour words such as words for pink and light blue (Russian goluboj, Polish niebieski) or light red (but darker than pink = Hungarian piros) in terms of these.

And this is how **Wierzbicka** comments on his approach:

What is particularly instructive about this passage is that the English color words are described here as “phenomenally basic” and are given in Roman type, whereas the Polish, Russian, and Hungarian color words are presented in italics. The implication is that in the case of Polish, Russian, and Hungarian, we are dealing with mere words, whereas in the case of English, no italics are needed, because we are dealing with reality itself. This being so, it is only logical that foreign words which don’t match “phenomenally basic” English words should be defined in terms of the latter. (Wierzbicka, 2014, p. 8)

What is her point of criticism? How is it related to the NSM approach?

5. Here is an abstract of a paper dealing with lexical typology titled *The semantic categories of CUTTING and BREAKING events: A crosslinguistic perspective*. (Majid, Bowerman, van Staden, & Boster, 2007). **Read it.**

Abstract

This special issue of Cognitive Linguistics explores the linguistic encoding of events of CUTTING and BREAKING. In this article we first introduce the project on which it is based by motivating the selection of this conceptual domain, presenting the methods of data collection used by all the investigators, and characterizing the language sample. We then present a new approach to examining cross-linguistic similarities and differences in semantic categorization. Applying statistical modelling to the descriptions of CUTTING and BREAKING events elicited from speakers of all the languages, we show that although there is cross-linguistic variation in the number of distinctions made and in the placement of category boundaries, these differences take place within a strongly constrained semantic space: across languages, there is a surprising degree of consensus on the partitioning of events in this domain. In closing, we compare our statistical approach with more conventional semantic analyses, and show how an extensional semantic typological approach like the one illustrated here can help illuminate the intentional distinctions made by languages.

Keywords: CUT and BREAK; separation events; verb semantics; categorization; extension; intension; typology; semantic map.

One of the starting points for this research was a study carried out by Clifton Pye in 1994 in three languages. The results are summarized in the table below: What does it suggest?

Compare it to Slovak. (Majid, Bowerman, van Staden, & Boster, 2007, p. 135)

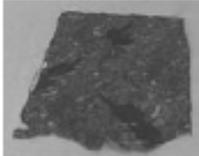
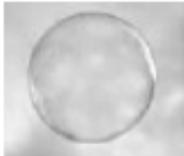
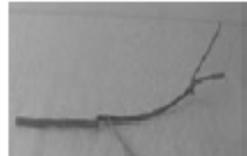
				
	cloth	bubble	plate	stick
English	<i>tear /rip</i>	<i>pop</i>	<i>break</i>	
Garifuna	<i>teiriguana</i>	<i>bowguana</i>		<i>halaguana</i>
Mandarin	<i>noŋ4-puo4</i>			<i>noŋ4-duan4</i>

Table 20 Comparison of C&B verbs in English, Garifuna and Mandarin (Pye 1994)

6. The following explications are given by Bromhead (Bromhead, 2011) for *river*:

The use of the word river is illustrated in the following examples:

- (1) . . . castles overlook the river's twisting course. . . (Collins Wordbanks British Ephemera)
- (2) . . . fortresses had been built along the lengths of the rivers. . . (Collins Wordbanks British Books)
- (3) The Zaire is the seventh longest river on earth. . . (Collins Wordbanks British Books)
- (4) . . .walking by the riverside with honeyed breezes blowing over the Shannon. . . (Collins Wordbanks British Books)
- (5) The rivers flow to the sea. . . (Collins Wordbanks US Books)
- (6) . . .they were crossing the great river on a creaking, paddle-driven ferry. (Collins Wordbanks British Books)

A river is a place where there is water, it is long in shape, and the water in it moves. The following explication [A] captures the features of the meaning of river:

[A] a river

- a. a place of one kind
- b. there is a lot of water [m] in places of this kind
- c. these places are long [m] places
- d. places of this kind have two sides [m]
- e. places of this kind are big places
- f. at many times when someone is on one side of a place of this kind if this someone wants to be on the other side, this someone can't be on the other side after a very short time
- g. the water [m] in places of this kind is always moving

- h. when someone is somewhere on one side of a place of this kind, this someone can think like this:
 “some time before this, this water [m] was in a place far from this place here
 some time after this, this water [m] will be in another place far from this place here”

River can be compared to stream which is smaller and has less water, and this water moves quickly. Some examples of stream

- (7) . . .before noon he came to a small stream. . . (Collins Wordbanks US books)
 (8) Long ago, a group of people lived along a stream. (Collins Wordbanks US Spoken)
 (9) . . . pretty camp spots beside mountain streams (Collins Wordbanks Oznews)

[B] a stream

- a. a place of one kind
 b. there is water [m] in places of this kind
 c. places of this kind are long [m] places
 d. these places have two sides [m]
 e. these places are not big places
 f. the water [m] in these places is always moving
 g. this water [m] moves quickly [m]
 h. when someone is somewhere on one side of a place of this kind, this someone can think like this:
 “some time before this, this water [m] was in a place near this place here,
 some time after this, this water [m] will be in another place near this place here”

Creek is exemplified by the following sentences:

- (11) The property also includes. . . fruit trees and a permanent creek. (Collins Wordbanks OzNews)
 (12) In the wet season the region becomes a huge catchment, filling the many usually dry creeks that flow from the higher country to the coast. (Collins Wordbanks OzNews)
 (13) A man died after his car ran off the road and into a creek bed. . . (Collins Wordbanks OzNews)

What is its explication? Fill in the missing words in the explication [C]. Choose from the following: *think, water, moving, side, big, long.*

[C] a creek

- a. a place of one kind
 b. often there is [m] in parts of places of this kind
 c. sometimes there is [m] in all parts of places of this kind
 d. these places are [m] places
 e. these places have two [m]
 f. these places are not places
 g. often the water [m] in these places is
 h. often when someone is somewhere on one side of a place of this kind, this someone can think like this:
 “some time before this, this water [m] was in a place near this place here,
 some time after this, this water [m] will be in another place near this place here”

The following explication is the explication of *karu* (Pitjantjatjara/Yankunytjatjara).

[D] *karu* ('creek bed, creek')

- a. a place of one kind
- b. sometimes there is water[m] (*kapi/mina*[m]) in places of this kind
- c. these places are long[m] (*wara*[m]) places
- d. these places have two sides [m]
- e. the ground[m] (*manta/pana* [m]) in a place of this kind is below the ground[m] on both sides of a place of this kind
- f. it is not like the ground[m] in other places, it is soft[m] (*tjula*[m])
- g. when there is water[m] in places of this kind, sometimes this water[m] is moving
- h. often there is water[m] in some places below the ground[m] in places of this kind

What is the main difference between *karu* and *creek*? Does *karu* exactly match the meaning of *creek*? Is the English translation *creek* correct?

7. Read the following explications (Wierzbicka, 2014, pp. 38-39):

wakatha ("a man's sister")

a man can say about a woman: "this is my *wakatha*"

if he can think about her like this:

"her mother is my mother, her father is my father"

yakukathu ("older sibling, same sex, female ego")

a woman (e.g. Margaret) can say about another woman (e.g. Elizabeth): "this is my *yakukathu*"

if she can think about her like this:

"her mother is my mother, her father is my father

She was born before I was born"

a child can say the same about another child

if after some time both these children can be women

duujinda ("younger sibling of same sex")

a woman can say about another woman (e.g. Margaret): "this is my *duujinda*"

if she can think about her like this:

"her mother is my mother, her father is my father

she was born after I was born"

a child can say the same about another child

if after some time both these children can be women.

a man can say about another man: "this is my *duujinda*"

if he can think about him like this:

"his mother is my mother, his father is my father

he was born after I was born"

a child can say the same about another child

if after some time both these children can be men.

thabuju ("older sibling, same sex, male ego")

a man can say about another man: "this is my *thabuju*"

if he can think about him like this:

"his mother is my mother, his father is my father

he was born before I was born”
a child can say the same about another child
if after some time both these children can be men.

Now go back to Chapter 1 to the Task section and use Figure 1 which illustrates the family tree of the British Royal Family. Use the explications and describe the family relationships, e.g. between Charles and Anna, Margaret and Elizabeth, Andrew, Edward and Charles etc.

Summary

What are the areas of study in lexical typology?

Describe the semasiological approach to lexical typology. Exemplify it.

Describe the onomasiological approach to lexical typology. Exemplify it.

What is understood by the lexicon-grammar borderline?

Describe NSM.

Areal typology

What is borrowing?

What is language contact?

What is the difference between bilingualism and diglossia?

What are sociolinguistic factors?

In the opening chapter to this textbook three reasons for similarities (and differences) between languages were mentioned.

First, languages share some properties because they are of the same origin. For example, the word for male parent in Slovak is *otec*. A similar phonetic make-up can be found in Slovene (*oče*), Russian (*omeu*), Croatian (*otac*), and Czech (*otec*). The reason for this similarity is obvious – the languages are of the same origin – they belong to the Slavic branch of the Indo-European family. A common origin (Germanic branch of the Indo-European family) explains similarity of the English *father* and the German *Vater*.

Second, languages are similar because they share a common cultural and social background. A shared cultural background can be exemplified by pronouns used to express politeness. In Slovak and French the second person singular (*ty, tu*) is used when speaking to friends and family members. However, in formal situations the pronoun used is in the second person plural (*Vy, vous*). The same situation can be found in Armenian, another Indo-European language but from the Armenian genus. However, there is no politeness distinction in West-Papuan Abun or Chibchan Bribri.

The third reason for similarities between languages is their contact. Let's take into consideration the word *pepper* by which I mean a powder of blackish colour that is used to add a hot taste to food. Compare the following languages:

Czech	Basque	Swahili	Maori
<i>pepř</i>	<i>piperra</i>	<i>pilipili</i>	<i>pepa</i>

Table 21 Words for pepper in Czech, Basque, Swahili and Maori

These languages are obviously genetically unrelated and still the forms for the meaning *pepper* display similarities. The country of origin of black pepper is south India. The origin of the word can be found in Sanskrit *pippali*. The merchants from India distributed to other countries not only spice but also its name. In this way languages got in touch, in contact.

Like shared origin and shared cultural background, **language contact**, too, can cause similarities between languages. Each contact needs some space. If this space is understood as a particular

geographical area, language contact can cause similarities between languages that are genetically unrelated. If linguistic convergence is observed in a larger group of languages sharing a common territory, the discussion of areal linguistics can be commenced.

Areal typology classifies languages into **areal types**, **Sprachbunds**, or **linguistic areas** according to shared properties, features, and typological parameters. This definition is very simple but, as is usually the case, simple definitions require complicated explanations.

According to Dryer (1989, p. 261), a linguistic area is an area in which “at least one linguistic property is shared more often than elsewhere in the world to an extent which is unlikely to be due to chance, but which is probably due either to contact or remote genetic relationships.” Enfield’s (2005, p. 190) more explicit definition accounts for a linguistic area as “a geographical region in which neighbouring languages belonging to different language families show a significant set of structural properties in common, where the commonalities in structure are due to historical contact between speakers of the languages, and where the shared structural properties are not found in languages immediately outside the area (ideally where these include languages belonging to the same families as those spoken inside the area).” Unfortunately, this broader definition makes the issue even more complicated. Let us therefore discuss it step-by-step.

Geographical region, neighbouring languages

The terms ‘geographical region’ and ‘neighbouring languages’ introduce the natural science of geography into linguistics. The question is what geographical factors should be taken into consideration. Borders of states definitely do not coincide with borders of linguistic areas; the opposite is true. The country borders are usually regions of rich language contacts that very often result in common linguistic features. An important role is played by natural geographical borders like mountains and rivers. In one of his studies, Bickel (2006) devotes attention to linguistic areas of Eurasia to find out that the highest density of rare features is in mountainous areas – the Caucasus and Himalayas. On the other hand, languages spoken in the plains share many common linguistic features. Logically, languages separated by natural borders are not in frequent contact with other languages and they maintain their linguistic features. Natural borders as such could solve the problem of language borders but the question is to what extent natural borders reflect the isolation of languages in the period of globalization, including profound social, political, economic, and cultural changes in the world. One of the suggestions proposes to rely on cultural borders (Dryer, 1989) as speakers of various languages living in one area very often share cultural traits. However, “shared culture does not imply the existence of a linguistic area” (Aikhenvald, 2007, p. 14). To exemplify this claim, she refers to the Great Plains area in North America that is recognized as a cultural area but not as a linguistic area.

Obviously, the question of the borders of linguistic areas cannot be avoided by typologists. They reach for another geographic tool – maps. Usually two types of maps are used for the study of internal structure of linguistic areas – maps with **isoglosses** and maps with **isopleths**. Isoglosses can be found in works of French dialectologists. They are based on a list of linguistic features that define a particular linguistic area. The territorial scope of each feature is examined and marked on a map; consequently, points on the map are connected by lines – isoglosses. A bundle of isoglosses specifies the core of a *Sprachbund*.

An example of isoglosses comes from Laver (2002 p. 75) (who adapted it from the work by Peter Trudgill):

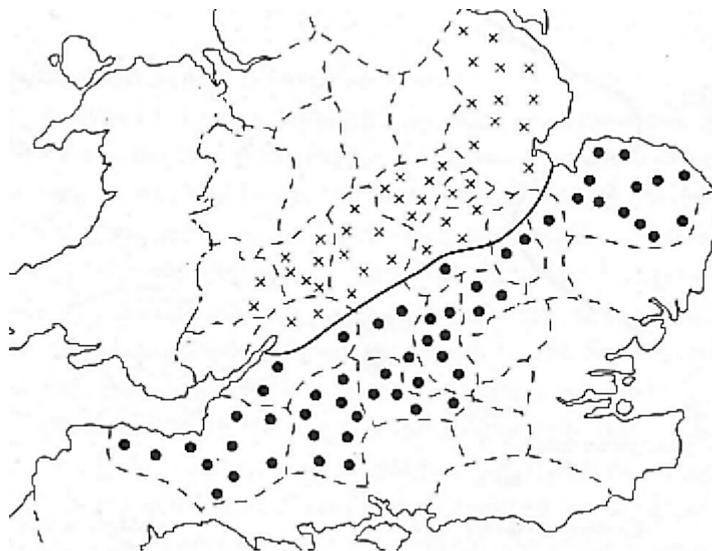


Figure 16 The /ʊ/-/ʌ/ isogloss boundary

The isogloss in map XY is the line that separates the dotted and the crossed areas. In the dotted area of southern England, the vowel in words like *cut* or *dust* is pronounced as [ʌ] while in the area with crosses the vowel is pronounced [ʊ].

Generally, isopleths are lines connecting points of equal value. In areal linguistics, similarly to isoglosses, they are based on a list of linguistic features shared by languages spoken in a particular linguistic area. If languages have the same number of properties, they belong to the same isopleth. The number of features in a given language determines the relative position of that language within the linguistic area. For example, the following map illustrates isopleth mapping of Africa. It works with 11 language features. There are two big centres. The first centre has nine and more properties and it is found in West Africa. It covers both Niger-Congo and Afro-Asiatic languages. Up to nine features were identified with languages in Cameroon-Central Africa. (Heine, 2010)

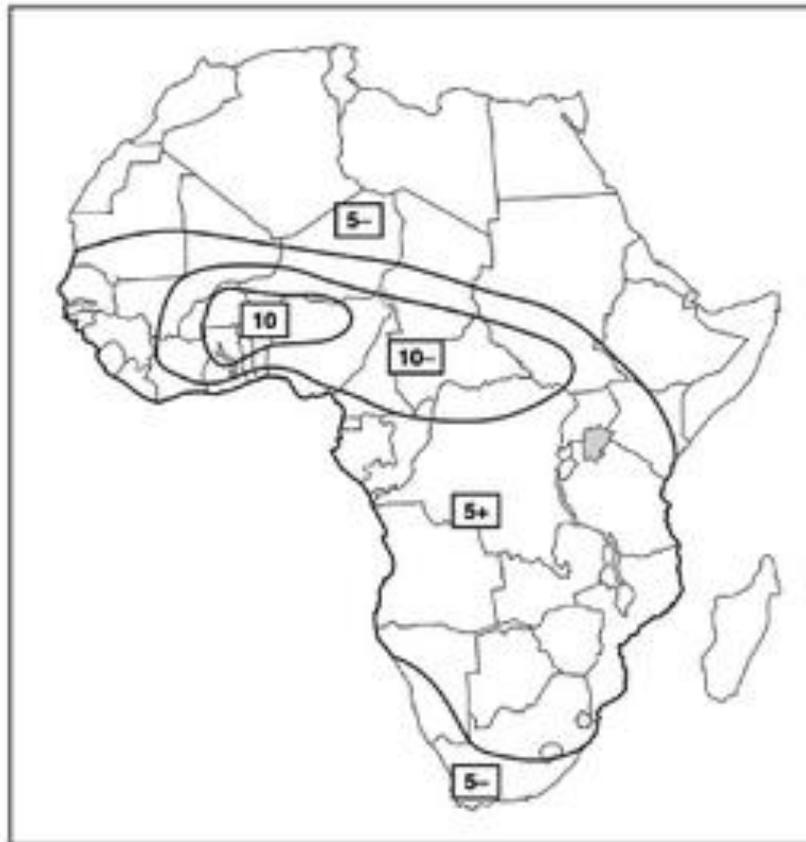


Figure 17 An isopleth sketch map of Africa based on 11 properties (99 language sample) (Heine, 2010 p. 6)

This begs the question as to whether it is indispensable and possible to have control over linguistic area borders. Each language as a dynamic system undergoes changes over time. These changes do not appear from day to day. They are gradual and unequal at individual structural levels in terms of degree, extent, and rapidity. However, they are in progress and can change the typological character of languages. As a result, they influence the shared features of languages and can move the borders of linguistic areas. That's why it may be proposed to eschew any effort to precisely define the borders of linguistic areas; instead of the Berlin-wall strategy let us conceive these areas as ones with open borders.

Language families and linguistic areas

The term 'language family' as defined by Enfield (2005, p. 190) projects the genealogical classification of languages onto areal typology. Languages belonging to one *Sprachbund* (linguistic area) may theoretically belong to different language families although, in reality, this kind of situation is rare: in the majority of linguistic areas languages are genetically related, or at least most of them. Languages in one linguistic area usually share two types of features – features due to language contact and features due to genetic relationships. The third type of shared features is represented by 'accidental'

features.¹⁵ These differ from the previous two types by not being anchored in historical development. Many of them are “ultimately due to the operation of linguistic universals of various kinds” (Thomason S. G., 2001, p. 100). For example, the fact that nearly every language has nouns and verbs does not suggest that they belong to the same linguistic area.

How many languages establish a Sprachbund? Linguistic areas are based on shared features. While the verb ‘to share’ suggests at least two participants, a common answer to the question of the number of languages is three and more. Campbell (2006) relates the concept of linguistic area to that of borrowing. He argues that there is no difference between “borrowing in general and areal linguistics in particular” (2006, p. 1). This assumption seems, however, to be problematic for two reasons. First, borrowing is a process and linguistic area is a situation. Second, and more important, while the process of borrowing is unidirectional, shared features within a linguistic area imply bidirectionality or, better, multidirectionality. By implication, these two notions cannot be used interchangeably. On the other hand, Campbell is right in saying that the process of borrowing is an inevitable condition for the existence of linguistic areas. Two sources of shared structural properties, history and language contact, have already been mentioned. A logical consequence of language contact is borrowing and multilingualism. This point of view is also held by Aikhenvald (2007). She adopts the definition of borrowing by Trask (2000) and understands it in its broad sense as ‘the transfer of linguistic features of any kind from one language to another as the result of contact’ (2007, p. 4). Borrowing is accompanied with linguistic diffusion defined as “the spread of a linguistic feature within a geographical area or as recurrent borrowing within a linguistic area” (Aikhenvald, 2007, p. 4). Linguistic diffusion can proceed from one source or from several sources. A long-term diffusion can result in the “layering of patterns and forms to such extent that genetic relationships are indiscernible” (Aikhenvald, 2007, p. 7). A case in point is the *Australian linguistic area*. Diffusion is enabled by various factors. Linguistic factors include pragmatic salience of a construction: the more pragmatically motivated the more diffusible; matching genres; tendency to achieve word-for-word and morpheme-by-morpheme translatability; frequency: the more frequent the category in one language, the more likely it is to diffuse into another; the more impact the category has on cultural conventions, the more diffusible it is expected to be; borrowing a practice may facilitate a set of linguistic expressions which correlate with it; existence of a perceivable ‘gap’ facilitates diffusion; etc.¹⁶

¹⁵ Cf. Thomason (2001).

¹⁶ Cf. Aikhenvald (2007).

Common structural properties

Enfield's term 'common structural properties', also called 'features' or 'variables', raises the question of quantity, i.e. 'How many features should be shared by Sprachbund languages?'; the question of structure, i.e., 'Should there be a sort of equal distribution of shared features by Sprachbund languages at all structural levels of language?'; as well as the question of proportionality, i.e., 'Should the shared features be proportionally distributed across individual Sprachbund languages? In terms of isoglosses, the intersection of features (that are represented in isoglosses) results in linguistic areas. Matras (2009, p. 268) assumes that a linguistic area "can be defined by a single feature that is areally patterned and crosses linguistic-genetic boundaries". Hence, the reply to the quantity question is one and more. The question of the distribution of shared features is answered by Thomason (2001, p. 101) "do all the features that are shared as a result of contact-induced change in a linguistic area have to be found in all the languages within the area? Here the answer is clearly no. If we answered yes, the number of linguistic areas in the world would immediately shrink from many to zero, because in all linguistic areas the distribution of shared features varies widely across the entire area." Similarly, Bickel (2007, p. 243) argues that "... hardly any typological variable, and only some combinations thereof, is evenly distributed in the world. Most distributions are subject to non-accidental geographic skewing." Bickel and Nichols (2006) identify several problems of structural features (they prefer the term 'variable') relevant to the definition of a linguistic area: there are no criteria for diagnostic variables; a language may be a recent immigrant to the area and the areal variables have not yet affected the language; the variables defining the area may be a motley set, etc. They suggest an opposite approach to structural features and name their approach *Predictive Areality Theory* (PAT). In their hypothetical linguistic area, any restrictions are based on a theory of population, language spread, and on information from other disciplines. Consequently, they test the hypothesis by "seeking statistically non-accidental signals" (Bickel & Nichols, 2006, p. 3). In this way, they employ multiple sciences in areal linguistics since population history requires archaeology, genetics, ecology, demography etc.

Sprachbund

It is generally acknowledged that the term Sprachbund was coined by Trubetzkoy (1928, p. 18).¹⁷

Gruppen, bestehend aus Sprachen, die eine grosse Ähnlichkeit in syntaktischer Hinsicht, eine Ähnlichkeit in der Grundsätzen des morphologischen Baus aufweisen, und eine grosse Anzahl gemeinsamer Kulturwörter bieten, manchmal auch äussere Ähnlichkeit im Bestande der Lautsysteme, – dabei aber keine systematische Lautentsprechungen, keine Übereinstimmung in der lautlichen Gestalt der morphologischen Elemente und keine gemeinsamen Elementarwörter besitzen, – solche Sprachgruppen nennen wir Sprachbünde.

¹⁷ For comparison of various definitions of linguistic area cf. Campbell (2006).

As pointed out by Emeneau (1956, p. 16), the German word *Sprachbund* was translated by Velten (1943) as 'linguistic area' and it spread quickly among linguists. This is how Emeneau (ibid) interprets the term:

This term 'linguistic area' may be defined as meaning an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families.

This widely applied definition has been completed by the issue of language contact. Thus, Sherzer (1959, p. 760) defines linguistic area as

...an area in which several linguistic traits are shared by languages of the area and furthermore, there is evidence (linguistic and non-linguistic) that contact between speakers of the languages contributed to the spread and/or retention of these traits and thereby to a certain degree of linguistic uniformity within the area.

Similarly, Haspelmath (2001) lays emphasis on genetically unrelated languages spoken in a common geographically defined territory, some features of which cannot be accounted for by their genetic origin:

A linguistic area can be recognized when a number of geographically contiguous languages share structural features which cannot be due to retention from a common protolanguage and which give these languages a profile that makes them stand out among the surrounding languages (2001, p. 1492).

Heine and Kuteva (2005, p. 172) distinguish three types of linguistic areas:

1. linguistic areas defined by the presence of a limited set of linguistic properties – Sprachbunds;
2. areas that are characterized by the fact that the languages concerned exhibit a high degree of mutual intertranslatability – metatypy;
3. areas that are the result of one and the same historical process, more specifically, of the same process of grammaticalization, even if there may be other properties in addition – grammaticalization areas.

Then, the term *Sprachbund* is justified whenever the following situation occurs (ibid):

- a. There are a number of languages spoken in one and the same general area.
- b. The languages share a set of linguistic features whose presence cannot be explained with reference to genetic relationship, drift, universal constraints on language structure and language development, or chance.
- c. This set of features is not found in languages outside the area.
- d. On account of (b.), the presence of these features must be the result of language contact.

As it follows from the preceding considerations, linguistic areas, *Sprachbunds*, owe their existence the phenomenon of language contact. Auwera (1998, p. 2) gives five reasons for shared features between languages. According to him

...the phenomenon α in one language and the phenomenon β in another language are similar or identical (are of a similar or identical type) because

- a. The two languages share another structural feature or they have two other structural features, again similar, and these other features are more basic and thus be argued to explain the initial identity or similarity – the structural explanation;
- b. They express a similar or identical meaning – the semantic explanation;
- c. They ultimately derive from processing principles (either of language production or understanding) – the functional or psycholinguistic explanation;
- d. They derive from the same phenomenon in the common ancestor language – the genetic explanation;
- e. They have arisen in a language contact situation with borrowing or calquing – the areal explanation.

Whatever criteria are applied, the question remains how many linguistic areas there are in the world. Linguistic areas and their boundaries are closely connected to extra-linguistic factors such as geography and culture. It is a well-known fact that extra-linguistic reality is a dynamic phenomenon, characterized by change. By implication, linguistic areas can and do change, too.

For obvious reasons it is not possible to give a definitive list of linguistic areas as can be done with language families and genera. Even though WALS and Ethnologue databases divide the world into macro-areas, those are not *Sprachbunds*. WALS, for example, divides the world into following macro-areas (the number in the brackets specifies the number of languages spoken in the given macro-area): Australia (177), Africa (607), Eurasia (660), North America (397), Papunesia (558), South America (257). Similarly, relying exclusively on geographical criteria, Ethnologue divides the world into Africa, Pacific, Americas, Asia and Europe. Each macro-area is further divided into smaller regions, e.g. Europe into Northern, Southern, Western and Eastern.



Figure 18 Macro-areas according to Ethnologue



Figure 19 Areas of Europe according to Ethnologue

It is obvious that WALS relies on geographical criterion, too. As a consequence, some languages can be assigned to 2 macro-areas, e.g. Avar (Nakh-Daghestanian) is matched with both Eurasia and Asia; Arabic (Afro-Asiatic) is assigned to Africa and Asia etc. Besides these macro-areas there are also smaller linguistic territories that are usually named after a geographical region they cover. So we have, for example, **The Balkan Sprachbund** and **The Pacific Northwest**. Each Sprachbund is determined by the member languages and features that are distributed among the languages.

So, for example, the Balkan Sprachbund is one of the first linguistic areas to have been identified. Its typological features were suggested by Trubetzkoy as early as 1928 (Friedman V. A., 2000). Its major languages belong to various language genera of the Indo-European language family. Rumanian is a Romance language, Balkan Romani is Indic; Bulgarian, Macedonian, and south-eastern dialects of Serbian are all Slavic languages; Albanian and Greek belong to their own genera. In addition, a crucial part of this Sprachbund is dialects of Turkish that are spoken in the Balkan. The Turkish language is a member of the Altaic language family.

Thomason (Thomason, 2000) lists the following features of the Balkan Sprachbund – the so-called **Balkanisms**: many Turkish and Greek loanwords in (other) Balkan languages; the presence of a high or mid central vowel, vowel harmony, the partial or total loss of the infinitive, postposed articles, pleonastic object markers, a merger of the dative and genitive cases, a future construction formed with the verb *want*, and a perfective construction formed with the verb *have*. Less distributed features are, e.g. the Slavic diminutive suffix *-ica* (in Greek, Rumanian, and Albanian), the plural suffix borrowed

into Rumanian from Greek, replacement of dative feminine pronouns with dative masculine pronouns in Macedonian as a result of Albanian influence, and the vocative case in Rumanian as a result of Slavic influence.

The Pacific Northwest is one of the richest linguistic areas of the world. It covers the region of north western US states Washington and Oregon and British Columbia, together with limited areas farther inland. The major language families are Salishan, Wakashan and Chimakuan. Some languages of the Athabaskan language family and other neighbouring languages (many of them isolates) share some of the areal features. Typical features include richly developed consonantal systems, predominantly polysynthetic type of languages, complex morphophonemics, reduplication, aspectual rather than tense distinctions, free word order.¹⁸

The Sepik River Basin, The Ethiopian highlands, The Baltic linguistic area, and South Asia are other examples of linguistic areas.

Standard Average European

It should also be noted that linguistic area is not the only term used in English to denote the phenomenon in question. Linguistic league (cf. (Friedman V. A., 2000), linguistic union (Du Nay, 1977), convergence area (Weinreich, 1953) are used, too. Boas, famous for his description of American Indian languages, spoke about acculturation and absorption (Boas F. , 1940). His follower, Benjamin Lee Whorf (1956) came up with the notion that has been widely discussed recently – the notion of *Standard Average European*. He studied Indian languages spoken in the USA and he put them in contrast to languages spoken in Europe that he “lumped ... into one group called SAE, or ‘Standard Average European’” (1956, p. 138). In recent decades this notion gained on attraction especially in connection with the idea of united Europe. The most intriguing questions are the borders of SAE and the catalogue of features that specify SAE languages. The questions are interrelated – the features under study influence the determination of the borders of the SAE territory. First of all, the nature of features is important. Décsy (1973), for example, relies on social and cultural criteria:

Zone	Languages
SAE	German, French, English, Italian, Russian
The Viking	Danish, Norwegian, Icelandic, Faeroese, Irish, Scottish-Gaelic, Welsh, Breton, Swedish, Lapp, Finnish, Vepsian
The Littoral	Frisian, Dutch, Basque, Spanish, Portuguese, Maltese
The Peipus	Estonian, Votic, Livonian, Latvian
The Rokytno	Polish, Lithuanian, Belorussian, Ukrainian, Kashubian
The Danube	Czech, Slovak, Hungarian, Slovenian, Serbo-Croatian (Serbian, Croatian, Bosnian)
The Balkan	Rumanian, Moldovan, Bulgarian, Macedonian, Albanian, Greek, Turkish

¹⁸ The literature on this topic is very rich, cf. for example (Nichols, 1992).

The Kama	Chuvash, Cheremis (Mari), Votyak, Mordvin, Ziryau (Komi), Yurak (Nenets), Kalmyk
Isolates	Luxemburgish, Romansch, Sorbian, Gagauz
Diaspora Languages	Yiddish, Ladino, Karaim, Romani (Gypsy), Armenian

Table 22 SAE according to Décsy

Haspelmath (1998), however, uses exclusively linguistic features. He compares the presence and absence of features like definite and indefinite articles, relative clauses with relative pronouns, ‘have’ perfect, nominative experiencers, participial passive etc. His classification of SAE languages is given in Table 23:

Zone	Languages
Nucleus	Dutch, German, French, northern Italian dialects
Core	Other Germanic and Romance languages, West and South Slavic and Balkan languages
Periphery	East Slavic, Baltic, Balto-Finnic, Hungarian, perhaps Basque, Maltese, Armenian, Georgian
clearly outside	Celtic languages, Turkic, Abkhaz-Adyghean, Nakh-Daghestan, eastern Uralic

Table 23 SAE according to Haspelmath

As both tables suggest, there has been an attempt to divide the territory of SAE into zones. While Décsy applies geographical names, Haspelmath divides the SAE languages into nucleus, core, periphery and languages outside the SAE. Auwera’s (1998) approach is a combination of both – he analyses linguistic features and applies geographical names to the individual zones:

Zone	Languages
Charlemagne Sprachbund	French, Italian, German, Dutch, Polish
Balkan Sprachbund	Rumanian, Albanian, Bulgarian, Greek
Northern around Danish	Danish, English, Norwegian, Faroese

Table 24 SAE according to Auwera

If we compare the tables above, it is obvious that French, German and Italian are identified as the central area of SAE languages. Interesting is the position of Slavic languages – Haspelmath excludes them from the nucleus; Décsy and Auwera list one Slavic language each – Russian and Polish, respectively.

However, if a different feature is chosen for analysis, there is a striking change of the situation. This claim can be exemplified by comparison of the previous analyses to research concerning the presence of diminutives and augmentatives in the languages of the world, with special attention paid to the SAE languages (cf. Körtvélyessy 2015). The methodology applied in the analyses differs from Haspelmath’s (and Auwera’s) but the results are comparable. On the basis of the so-called saturation value¹⁹, the

¹⁹ Cf. Körtvélyessy (2015). Generally speaking, saturation value is understood as a numerical representation of the structural richness of productive word-formation systems, processes, types and rules used to form new complex words.

SAE languages are divided into colour zones where the brown colour identifies what Haspelmath labels as nucleus:

zone	Languages
brown	Italian, Occitan, Slovak, Slovene, Belorussian, Czech, Portuguese, Polish, Provençal, Greek, Spanish, Upper Sorbian, Russian, Lower Sorbian, Serbian
green	French, Romansch, Bulgarian, Kashubian, Ukrainian, Zazaki, Turkish, Ladino, Roman, Lithuanian, Macedonian, German, Basque, Catalan, Dutch, Norwegian, Tatar, Breton
blue	Ossetic, Hungarian, Mansi, Croatian, Scottish Gaelic, Latvian, Maltese, English, Icelandic, Votic, Armenian, Danish, Estonian, Finnish, Karelian, Welsh, Mordvin, Saami, Udmurt
red	Albanian, Faroese, Irish, Luxemburgish, Udi, Mari Meadow, Gagauz, Ingush, Adyghe, Kabardian
yellow	Khvarsi, Dargwa, Aghul, Archi, Lezgian, Tabassaran, Komi-Permyak, Khanty, Ubykh

Table 25 SAE according to Kortvelyessy (2015)

If these two tables are compared, the brown area overlaps to some extent with Haspelmath's nucleus. The biggest difference is the absence of Germanic languages in the nucleus and the central position of Slavic languages in it. The green area is comparable with Haspelmath's core. However, in addition to the Germanic, Slavic and Balkan languages there are also Altaic, Celtic and Baltic languages and Basque, i.e., languages that are 'clearly outside' according to Haspelmath. On the other hand, in Haspelmath 'clearly outside' there are languages of the Nakh-Daghestan family and Eastern Uralic languages – the languages included in the yellow area (languages without diminutives and augmentatives).

To conclude, SAE borders depend on the analysed language features. They are not fixed and can be more precisely identified by comparison of the results of the analysis of all the significant language features ranging over individual language levels – phonology, morphology, syntax, lexicology.

Tasks

- 1. What is a *Sprachbund*? Compare it to the definitions of a linguistic area mentioned in the text above.**
- 2. According to one definition, a linguistic area should contain a group of at least three languages. Why are two languages not sufficient? You can find the answer in the following paragraph (Thomason, 2000 p. 1):**

.... Why not just two languages in contact? The most obvious reason is that subsuming two-language contact situations under the rubric 'linguistic area' would mean that almost every contact situation in the world that involves significant structural interference would be a linguistic area; and although there are important similarities between interference in two-language contact situations and interference in more complex contact situations,

there are also important differences. Structural interference in many or most Sprachbünde is multidirectional, for instance, while structural interference in many or most two-language contacts is unidirectional: so, for instance, it is clear that the changes that formed the network of shared features in the Balkan Sprachbund did not all originate in the same language, while structural interference in the Romani-Russian contact situation is all from Romani to Russian, not vice versa (at least as far as Russian as a whole is concerned). But some linguistic areas, such as the Ethiopian highlands, seem to have unidirectional interference resulting from language shift (see discussion below); and in some two-language contacts, such as Uzbek and Tadjik in the former USSR, structural interference can be found in both languages, though it didn't necessarily happen simultaneously or throughout the contact region (see Comrie 1981: 51, 163; for a broader view of Turkic-Iranian contacts, see e.g. Johanson 1992, 1998). From a historical linguist's viewpoint, perhaps the major reason for considering two-language contacts separately from Sprachbünde is that in the great majority of cases the source of a shared feature is easier to determine when only two languages are involved.

3. **How many features are present in one linguistic area? Is it possible to answer this question?**
4. **Which languages belong to the Sepik River Basin and The Ethiopian highlands language areas?**
5. **Read the following text. What kind of text is it – what was its purpose? What linguistic areas are mentioned? What is EUROTYP? What is the difference between areal linguistics and typology in areal context? What languages create the Circum-Baltic area?**

On two recent publications on areal linguistics

Östen Dahl

It would be fair to say that interest in the phenomenon of linguistic areas ('areal linguistics' or 'areal typology') has been on the rise during the past decade. Whereas earlier, the cases of the Balkans and Southwest Asia were classic but rather isolated textbook examples of language convergence, during the 1990s new discussions of linguistic areas appeared, focusing on Mesoamerica, East Asia, Siberia, the Ethiopian Highlands, and Anatolia, among others, and the notion of linguistic convergence has begun to occupy a centre stage position in discussions surrounding the classification of Amazonian, Australian, and Papuan languages. In the specifically European context, convergence entered the typological discussion in connection with the EUROTYP project of the early 1990s and the formulation of observations on the geographical clustering of typological features. The two collections under review are the best representatives of this latter development, engaging contributors among both EUROTYP veterans and other typologists with expertise in one or more of the relevant languages.

The Circum-Baltic collection (edited by Östen Dahl and Maria Koptjvskaja-Tamm) takes a parallel approach to what it defines as two separate agendas: Areal linguistics, which is the study of individual languages in an area, as well as of the historical connections among them: and typology in an areal context, which is the study of grammatical phenomena within a particular area. This is reflected in separation of the two volumes of the collection. Volume 1 deals with the languages of the area. It begins

with surveys of the principal languages – Lithuanian and Latvian (both Laimode Balode and Axel Holvoet), Swedish (Anna-Charlotte Rendahl), Russian (Valeriy Ekmonas) and the Finnic languages (Johanna Laakso) – focusing in particular on dialect variation. Two historical studies follow: One on the origin of the Scandinavian languages (Östen Dahl), the other on Baltic influence on Finnic languages (Lars-Gunnar Larsson). The third section, devoted to ‘Contact phenomena in minor Circum-Baltic languages’, departs from the format of the other language-specific chapters, in that the issues dealt with are more specific and more selective, general background about the languages is not always provided, the pre-theoretical approach that is typical of the first section is not always maintained, and we find various degrees of exhaustiveness even in relation to the discussion of contact phenomena. The intention is obviously to give consideration within the collection to lesser-known and smaller languages of the region: Karelian (Finnic), Karaim (Turkic), Yiddish (West Germanic), Romani (Indo-Aryan), and the Northwest Central Russian dialect.

6. Slovak and Hungarian are genetically distinct languages. Can you explain the following lexical similarities between them?

Slovak: *klobúk, gombík, streda, vedro, lopata*

Hungarian: *kalap, gomb, szerda, veder, lapát*

7. Find an explanation for the following similarities and difference at the level of lexicon:

English	Yoruba	Italian	Estonian	Maori	Romanian	Swahili	Azerbaijan	Slovak	Albanian
tea	tii	tè	tee	tea	ceai	chai	çay	čaj	çaj

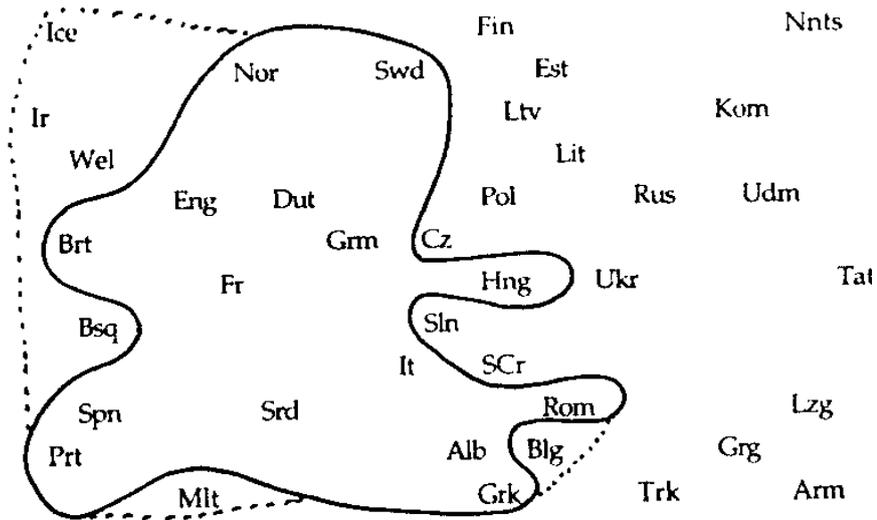
Table 26 Tea

8. The map in WALS, chapter 138, suggests that the words for *tea* are derived either from Sinitic *cha* or from Min Nan Chinese *te*. What about Polish *herbata* or Lithuanian *arbata*?

9. What do you think is the reason for similarities in the word for *sugar* in various languages of the world? For example, we have *sugar* in English, *Zucker* in German, *cukor* in Hungarian and Slovak, *sukari* in Swahili, *suga* in Yoruba etc. Why does not such a similarity work for *salt*? While it is *salt* in English, *Saltz* in German, *soľ* in Slovak, *sól* in Polish and *sel* in French, it is *iyọ* in Yoruba, *chumvi* in Swahili and *tuz* in Turkish.

10. Go to ethnologue.com. Find the territory of Americas. How is it further divided? What kind of information can you find there?

11. The map below illustrates the usage of the definite and indefinite articles in the languages of Europe. Comment on it. What are the lines on the map called?



..... definite and indefinite article present

_____ only definite article present

Figure 20 Definite and indefinite article (Haspelmath 2001)

Besides the definite and indefinite articles, Haspelmath in his work (2001) analysed eight more features (cf. the text above) and the results were combined in the following map. Comment on it:

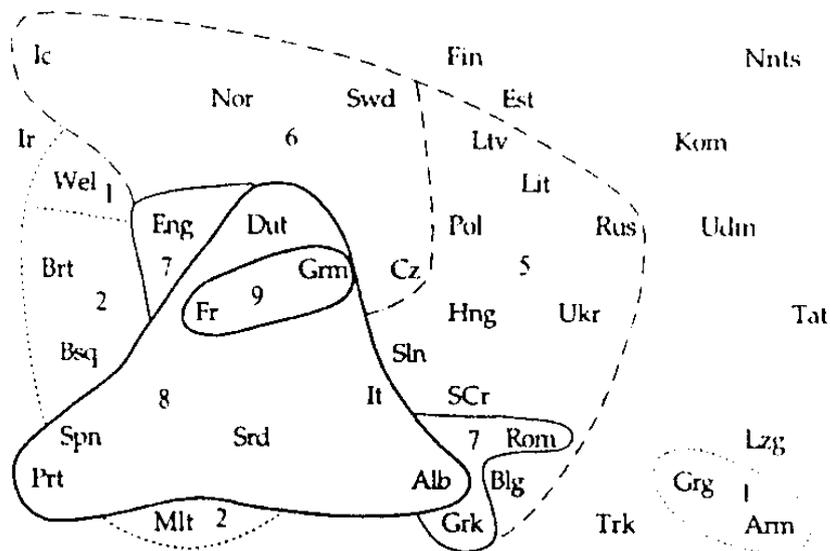


Figure 21 A cluster map combining nine features (Haspelmath 2001)

Summary

What is a Sprachbund?

What are the reasons for the existence of a Sprachbund?

How many languages are in a Sprachbund?

How many features are shared by languages in one Sprachbund?

How many Sprachbunds are there?

What is SAE? What are its borders? What features are shared by SAE languages?

Language universals

What is Universal Grammar?

What are the basic principles of transformational and generative grammar?

What do you know about Noam Chomsky?

What are the basic properties of human language?

What are the basic principles of the Prague Linguistic Circle?

What do you remember about Trubetzkoy and Jakobson?

What is meant by synchrony and diachrony?

".. we do not want to invent language universals but to discover them. How to discover them is not so obvious."

Hockett (1963, p. 1)

In the first chapter five titles of books dealing with language universals were discussed. Two of them (Comrie, Haspelmath), explicitly state that the book is not only about language typology but also about language universals. But also the remaining three books clearly suggest that the concept of language universals is unavoidable. It seems that language typology and language universals are inseparable and study of one field goes hand in hand with the other. While the previous chapters focused on various aspects of language typology, in the following lines attention will therefore be paid to the topic of universals.

Quest for the right label

Literature on language universals uses three synonymous labels for the field of research discussed in this chapter: language universals, linguistic universals and universals of language. The term used depends on the approach applied. In the following, various approaches to universals will be discussed. Since terminology is crucial to the comprehension of the issues in question, the chapter starts with clarification of the individual terms and concepts.

It is generally assumed that universals are the ultimate aim of linguistic studies (e.g. (Aitchison, 2003, p. 4). However, there are various ways how to reach this aim; different authors employ different terminology and methodology.

Who uses what?

As it will be demonstrated, there are two main streams in the study of language universals. They could be named after the main representatives, J. **Greenberg** and N. **Chomsky**. Representatives of the Greenbergian stream use the term language universals while Chomsky's followers work with linguistic universals. Since these terms are not infrequently used as synonyms let us clarify their respective scopes.

Language universals

Hockett (1963, p. 1) defines language universal as a "feature or property shared by all languages, or by all language. The assertion of a (putative) language universal is a generalization about language". Similarly, Greenberg in his famous *Memorandum Concerning Language Universals* (1963, p. 255) states:

"Language universals are by their very nature summary statements about characteristics or tendencies shared by all human speakers. As such they constitute the most general laws of a science of linguistics (as contrasted with a method and a set of specific descriptive results). Further, since language is at once both an aspect of individual behaviour and an aspect of human culture, its universals provide both the major point of contact with underlying psychological principles (psycholinguistics) and the major source of implications for human culture in general (ethnolinguistics)."

This definition introduces into the discussion on language universals the intricate questions of human behaviour and culture. As Greenberg himself points out (1963, p. 262) the study of universals leads to "a whole series of empirical generalizations about language behavior". Furthermore, he identifies the problem of universals in the study of human language and in human culture (1976, p. 9): "The search for universals, therefore, coincides on this view with the search for laws of human behavior, in the present context more specifically those of linguistic behavior" (ibid). Therefore, "... any generalization about spoken language is also a hypothesis about human cultural universals" (Hockett, 1963, p. 11). Thus, the search for language universals is connected to the study of human culture and language behavior.

Universals of human language

The word 'human' in the title of this section is used to contrast natural human languages with languages of animals. Based on this distinction Hockett (1963) introduces an idea of features of human language also known as **design features of language**. The design features of language are reflected in universals of human language, e.g., every human community has a language; every human language has a tradition, but also changes over time; every language can express unrestricted meanings; every language has duality of patterning etc.

Linguistic universals

The term is preferably used by followers of Chomsky. Since Universal Grammar is innate for them, they search for linguistic universals in an individual speaker-hearer. In comparison to Greenberg who interrelates language and culture, Chomskians' search for linguistic universals in Universal Grammar focuses on various aspects of neurology and biology.

However, there is also a different understanding of the term 'linguistic universals'. Plank and Lahiri (2009, p. 31) point out that "it can be deceptive to axiomatically equate 'linguistic universals' with 'universals of language'. They maintain (2009, p. 32) that

"...it is not only difficult but impossible to prove universal claims – linguistic universals as well as cognitive or phonetic universals in which they may be grounded – which are meant to be true of the domain of ALL human languages EVER; empirically true, that is, rather than true by definition of 'language'".

They approach linguistic universals from the point of view of the specific aspect of language. This approach naturally follows from the structure of Plank's **Universals Archive** (<http://typo.uni-konstanz.de/archive/intro/index.php>) where a universal domain can be syntax, morphology, word-formation, stylistics, discourse, etc.

One of numerous interesting points in the Universals Archive is that concerning **rara**. Plank describes them as the other end of the scale of language universals. An extreme case of rarum is 'rarissimum'. It is "... a trait ... which is so uncommon across languages as not even to occur in all members of a single ... family or discussion area ... Diachronically speaking, a rarum is a trait which has only been retained, or only been innovated, in a few members of a single family or Sprachbund or of few of them" (Plank, 2000). Bickel and Nichols (2003, p. 3) distinguish two types of rara: 1. absolute rara that are found rarely across language families and 2. relative rara that are rare on a global scale but common within geographical area or a language family. Plank, Bickel and Nichols are not the only ones who see the importance of rara for a theory of universals. They are also mentioned by Cysow and Wohlgemut (2010) who call them 'unicale'.

In this short terminological introduction various problems of language universals, which will be elaborated on in this chapter, were briefly indicated. Furthermore, two poles of one scale were introduced – language universals and rara. The space between them is a continuum that evokes many questions – from deep conviction about the existence of universals through calling their existence into question up to absolute negativism.

Since Darwin language has been frequently compared to a living organism. Let's take advantage of this metaphor and compare language to the human body. Even though each human being is an individual, our anatomy is basically the same. Analogically, the skeleton of language is built by Hockett's design features. What creates the flesh, muscles and veins is influenced by many geographical, historical, cultural and social factors. In the same way as blood flows through veins and transports substances

necessary for life, these factors run through the body of a language. What makes human beings different from one another is their appearance and nature, both deeply rooted in genetics. Similarly, what makes languages different is not their skeleton, flesh and muscles but their nature. A combination of various factors results in similarities but also differences. Moreover, the body of language is dressed up and to find either resemblance or differences means to uncover many layers.

There are various research methods for uncovering similarities and distinctions among languages. Generally, they are employed by comparative and contrastive linguistics. While comparative linguistics searches for similarities among languages, the contrastive approach focuses on the differences. The comparative approach requires comparanda and one tertium comparationis. Tertium comparationis should be an integral part of language activity, but not identical with an individual language (Seiler, 2000, p. 29). Contrastive research very often results in the above-mentioned scepticism to the existence of language universals. This position is expressed by Evans and Levinson whose paper *The myth of language universals: Language diversity and its importance for cognitive science* (2009) triggered a frenetic discussion as they oppose many deeply rooted convictions about language/linguistic universals. They object mainly to the conclusions of the generative approach and label the claims of Universal Grammar as “empirically false, unfalsifiable, or misleading in that they refer to tendencies rather than strict universals” (2009, p. 429). They criticize the employment of Universal Grammar as an “empirically verified construct” (ibid) and regard Chomsky’s concept of Universal Grammar as mistaken based on “a set of substantial research findings about what all languages have in common” (2009, p. 430). Rather than similarities among languages they point out differences: “The crucial fact for understanding the place of language in human cognition is their diversity” (2009, p. 431). Their scepticism to language universals and their emphasis on language diversity is influenced by numbers – they state that we have information about 10% languages out of approximately 7000. Moreover, language death-rate is on average one language every two weeks. On top of this, there are about 100 isolates. Consequently, they conclude that “...nearly all generalizations about what is possible in human languages are based on a maximal 500 languages sample” (2009, p. 432).

Evans and Levinson’s paper represents the third possible approach to language universals. While Greenberg based his theory on comparison of languages and Chomsky on study of an individual’s Universal Grammar, Evans and Levinson begin their theory by stressing the importance of diversity among languages. These three general approaches are in accordance with the already mentioned continuum.

[Approaches to language universals](#)

A thorough and systematic study of languages started in 17th century. The concept of universal reason emerged with the scientific revolution and the Enlightenment. According to this concept “the general

takes precedence over the particular, the abstract over the concrete and the non-temporal over the historical” (Mairal & Gil, 2006, p. 2). The effort of philosophers, linguists and pedagogues (Descartes, Leibnitz, Locke, Diderot, and Rousseau) was aimed at creating new artificial languages and universal languages. It was Cartesian philosophy that introduced the comparison of languages. One of the basic premises of Cartesian philosophy was “the defence of innateness, or the belief that if objects in the real world are knowable, which they evidently are, it is because of the existence of innate ideas or conceptual structures that have not reached us by way of our senses or imagination, and which are not generalizations made by induction, or are even in need of empirical confirmation” (ibid). The conclusion was that innate ideas are universal. It was opposed by empiricism according to which all knowledge comes from perception and thus cannot be derived from innate principles. Contrary to this, rationalism (which strongly influenced the generative approach), insisted on the existence of innate ideas or, to use Leibniz’ term, truths of reason.

The 19th century saw the birth of comparative linguistics. Even though linguists devoted their attention to the origin of language (e.g. Herder: *Abhandlungen über den Ursprung der Sprache*. 1770), Sanskrit was ‘discovered’ and the rules of consonantal shifts were formulated (Grimm, Verner), languages were compared (Rask, Bopp) and structures of languages were studied (Humboldt), the real beginning of the study of language universals dates back to Greenberg and Chomsky. The dialogue between these two movements is one of the main characteristic features of linguistics of the 20th century. Both streams had their conferences and publications – Greenberg the *Dobbs Ferry Conference on Language Universals* (1964) and *Universals of Language*; Chomsky the conference *On Universals in Linguistic Theory* held at the University of Texas, Austin, 1967 and *Universals in Linguistic theory* edited by E. Bach and R.T. Harns.

Both conferences had their successors (e.g., the Language Universals Project at Stanford University 1967-76; a series of public lectures on universals in summer 1971) and they had a great impact on the development of ideas in Europe (a conference on typology and universals in Moscow 1966 organized by Uspenskij; XI International Congress of Linguists in Bologna in 1972; the Universalien project at the University of Köln directed by Hansjakob Seiler in 1972). Skipping a few decades, we are in the 2000’s and at the Cornell Symposium on Language Universals (Morten H. Christiansen, Chris Collins, and Shimon Edelman) that took place at the beginning of May 2004, as well as at the international congress held at the University of Bologna in January 2007 (organized by Sergio Scalise, Elisabetta Magni, Antonietta Bisetto) and the conferences on Typology and Universals in Word-formation in 2009, 2012, and 2015 (organized by P. Stekauer). Of course, a project with world significance at the Konstanz University directed by Frans Plank and Elena Filomonova must be listed here, too. The Universals Archive appeared on the Internet in 1999 and became an inspiration for further research.

When comparing the two approaches to the study of language universals, their background is worth mentioning. In Greenberg’s case it is his approach to language as a complex system, as well as theoretical contribution of the Prague School, especially that of Jakobson and Trubetzkoy. The participants of the Austin conference were influenced by the ideas of Noam Chomsky and his Universal Grammar. Interestingly enough, both Greenberg and Chomsky were inspired by Jakobson – while Greenberg found inspiration in Jakobson’s application of Trubetzkoy’s marked and unmarked categories to grammatical categories and semantics, Chomsky’s basic idea about the innate nature of universals comes from Jakobson’s observation of “similarities in children’s acquisition of phonology across languages and preferred patterns of phoneme distribution across languages” (Bybee, 2009, p. 17).

The chart below illustrates the two most important fields discussed at both conferences – the type of universals and the topics discussed.

topic/conference		Dobbs Ferry	Austin
Universals proposed		non-definitional	definitional
		statistical	non-statistical
		implicational	non-implicational
Topics discussed and their importance:	typology	+	-
	diachrony and language change	+	-
	archiving	+	-
	listing and numbering of universals	+	-
	sampling of data	+	+/-
	link to other disciplines	+	+/-
	acquisition of language	+/-	+
	focus on one language mainly	-	+
	innate knowledge of language universals	-	+

Table 27 Summary of two conferences on language universals

The typology of languages, “...the nongenetic, nonareal classification of languages as a valuable means to the discovery of universals” was emphasized by Hockett, Saporta and Greenberg and deemphasized (Ferguson, 1978, p. 11) by the conference in Austin. The reason for this controversy is obvious – while the Dobbs Ferry conference stressed the importance of comparison of as many languages as possible, participants of the Austin conference exemplified their statements basically using one language – English (the only exception was Fillmore comparing 20 languages). Following the principles of Universal Grammar, they understood linguistic universals as innate. By implication, an in-depth understanding of one language in depth seemed to be sufficient. Another logical consequence of this approach is the interest in language acquisition. Moreover, the importance of diachronic approach and language change for the study of language universals was one of the central topics at Dobbs Ferry; in Austin, it was only mentioned by Fillmore. While participants of the Dobbs Ferry conference pointed out the necessity of linking linguistics to other disciplines, especially anthropology and psychology, in Austin

the influence of philosophy and mathematics was visible. The significance of sampling methods, the archiving, listing and numbering of universals is another logical sequence of the typological approach. These issues were formulated and discussed in conference papers (e.g. Hockett and Greenberg). As mentioned by Ferguson (Ferguson, 1978, p. 16), the universals at Austin were presented so discursively that “if one asks exactly what substantive and formal universals were proposed in the papers, it is not always easy to determine the answer”.

The two movements should not, however, be viewed as competing. The opposite is true. As pointed out by Moravcsik (Körtvélyessy, Interview with Edith Moravcsik, 2016), they are entirely complementary since one represents the theoretical facet and the second the facts.

Types of universals

In the Greenbergian approach there are three basic principles underlying the typology of universals:

1. the type of language universals is determined by the language level they refer to – phonological, morphological, syntactic, semantic ...;
2. presence vs. non-presence of a universal in languages results in ABSOLUTE VS. STATISTICAL universals;
3. need of reference to another property and language – IMPLICATIONAL VS. UNCONDITIONAL universals.

A universal in the Greenbergian sense is an intersection of these three views, e.g. #Universal12 from the Universals Archive. Its domain is both morphology and syntax, it is of implicational type and statistical quality.

The Universals Archive

Number	12
Original	Consistent VO languages tend to be inflectional in their morphology.
Standardized	IF basic order is consistently VO, THEN the morphology tends to be flexive.
Formula	VO ⇒ flexion
Keywords	order, VO, flexion
Domain	morphology, syntax
Type	implication
Status	achronic
Quality	statistical
Basis	languages in Lehmann 1973
Source	W.P.Lehmann 1973 : 47
Counterexamples	—
Comments	By Frans Plank 03.08.2006, 09:49 See also . "Flexion" designates a morphological type, as opposed "inflection". Its hallmark is cumulative exponence (which is sometimes also called "fusion" after Sapir, though Sapir's notion of "fusion" is arguably not equivalent to cumulation: adjacent morphemes can be "fused" through regular phonology, or they can be morphologically cumulated in the first place).

Figure 22 #Universal 12 in The Universals Archive

Universals at language levels

As early as the conference in Dobbs Ferry, Hockett presented 10 'grammar' and 10 'phonological' universals. The starting point of his proposal was the comparison of human language to the language of animals that resulted in the universals of human language. Similarly, Greenberg listed 25 universals for the domains of syntax and morphology.

The present-day number (July 2016) of universals listed in the Universals Archive is given in the chart below²⁰. The most numerous domains are syntax and inflection, followed by phonology. The differences in numbers are striking.

domain	number of universals listed
syntax	1129
inflection	789
phonology	543
lexicon	158
morphology	157
semantics	142
prosodic phonology	62
word-formation	51
pragmatics	14
discourse	11

Table 28 Domains in the Universals Archive

²⁰ The total number of Universals in the Universals Archive is 2029.

Absolute vs Statistical Universals

Greenberg (1976, p. 9) maintains that “for a statement about language to be considered fully general it is sufficient that it has as its logical scope the set of all languages.” Absolute universals are exceptionless, “obvious universals, e.g. all languages have vowels, also those involving numerical limits, e.g. for all languages the number of phonemes is not fewer than 10 or more than 70; every language has at least two vowels” (Greenberg, Osgood, & Jenkins, 1963, p. 259). While Comrie (1981, p. 19) defines absolute universals as “those that are exceptionalness, and those that exist as tendencies, but do still have exceptions”. He understands tendencies as “statistical significant deviation from random patterning. An absolute universal, in this sense, is just the extreme case of deviation from random distribution.” (ibid)

Statistical universals are defined by Greenberg in the following way (1963): “For any language a certain characteristic (φ) has a greater probability than some other (frequently its own negative). This includes ‘near universals’ in extreme cases [...] For example, of the three devices of suffixing, prefixing and infixing, the probabilities are not random and are stated here in decreasing order.”

Statistical universals present one major problem. The number of cross-linguistic samples is usually very small in comparison to the total number of languages. Greenberg based his 45 universals presented at the Dobbs Ferry conference on comparison of merely 30 languages. The geographical distribution could also be called in question. That’s why some linguists doubt the existence of universals; e.g. Johanna Nichols asks (2008, p. 292): “If universals cannot be detected by statistical means and most structural analysis is framework-internal, how are we to know universals?” Also Hockett (1963, p. 1) was aware of these circumstances when he argued: “The assertion of a language universal must be founded on extrapolation as well as on empirical evidence” ... “We should rather formulate generalizations as hypotheses, to be tested as new empirical information becomes available”.

Greenberg (1963) himself realized this fact. When he assumes in the Memorandum that universals are of logical structure, he divides universals into universals which concern existence (unrestricted, implicational and restricted) and universals that concern probability (statistical universals, statistical correlations and universal frequency distributions). His understanding of statistical universals is mentioned above. Statistical correlations are described as follows: “If a language has a particular characteristic (φ) it has a significantly greater probability of possessing some other characteristic (Ψ) than if it does not possess (φ). E.g. usually a language with gender distinction in the second person singular also has its distinction in the third person singular but not vice versa. If this were without exception, we would have the implication: second person singular pronominal gender \rightarrow third person singular pronominal gender. There are languages in central Nigeria that have the distinction the second person, but not in the third. If the exceptions are genuine, then we have the following statistical

correlation: If a language has a pronominal gender in the second person singular, it has a greater probability of having this distinction in the third person singular than of not having it". Universal frequency distributions are based on the fact that a certain measurement may be applied to any language. "When this is so, it is possible that the results of each measurement over an adequate sample of languages will show a characteristic mean and standard deviation. Means, standard deviations, or other statistical measures derived from such distributions may be considered as universal facts about language." (ibid).

For illustration, the Universals Archive counts 1237 absolute and 827 statistical universals.

Implicational vs. Unconditional universals.

The following type of universals concerns existence of a language property. Unrestricted, unconditional universals are possessed by all languages "which are not merely definitional; that is, they are such that if a symbolic system did not possess them, we would still call it a language" (Greenberg, Osgood, & Jenkins, 1963, p. 258). For illustration, all languages have vowels; for all languages the number of phonemes is not fewer than 10 or more than 70. Universal implications always involve relationship between two characteristics. "If a language has a certain characteristic, (ϕ), it also has some other particular characteristic, (ψ), but not vice versa. That is, the presence of the second characteristics (ψ) does not imply the presence of the first one (ϕ)" (1963:259). If a language has a category of dual, it also has a category of plural but not necessarily vice versa. In the following, we express such relationships between predicates by an arrow, for example, dual \rightarrow plural. (1963:258). Cases of mutual implications between characteristics which are not universal are called restricted universals. "If any language has a particular nonuniversal characteristic ϕ , it also has ψ and vice versa." (1963:259) E.g. if a language has a lateral click, it always has a dental click and vice versa.

Greenberg's typology of universal illustrates the following scheme:

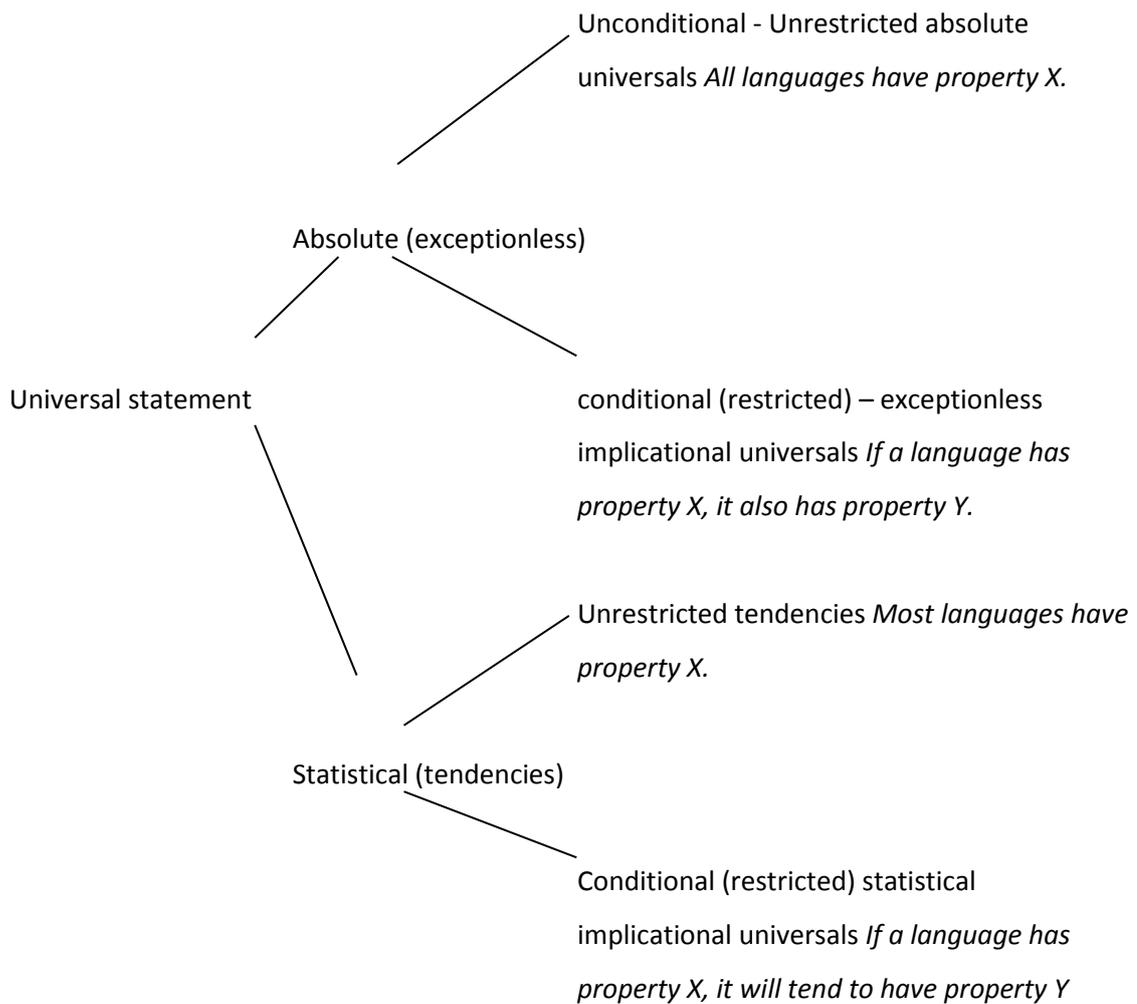


Figure 23 Greenberg's typology of language universals

Chomsky developed his typology of universals, too. His starting point is the innateness of Universal Grammar. Since his approach is confined to the analysis of a single language, any typology is, for obvious reasons, excluded. In his theory of syntax (1965) he assumes that “(t)he study of linguistic universals is the study of the properties of any generative grammar for a natural language” (1965, p. 28). In accordance with the theory of deep and surface structures, he suggests a classification of universals into **formal universals** and **substantive universals**. “A theory of substantive universals claims that items of a particular kind in any language must be drawn from a fixed class of items”. (ibid). This can be illustrated with Jacobson’s distinctive features and, generally, with all universals proposed by Greenberg. On the other hand, formal universals are of more abstract sort. They are properties of all natural languages, e.g. colour words of any language must subdivide the colour spectrum into continuous segments.

Synchrony vs. diachrony, language change, explanation of universals

As already mentioned, while Chomsky rejects the topic of diachrony in his considerations about language universals, Greenberg assigns it a crucial role. In his *Memorandum* he states that synchronic and diachronic regularities are obviously interrelated. He also speaks about universals of change that have important psycholinguistic implications (ibid). They are probabilistic. The general rule is as follows: "For all (x) and (y) where (x) is an earlier and (y) a later stage of the same language". Greenberg stresses the fact that both synchronic and diachronic linguistics are "seen to complete each other in that neither can be fully understood without the other" (1963). Some other conference participants, for example Hoenigswald (1963), raised the question of diachronic universals and the role of language change.

Later on, Chomsky's followers, too, discovered the importance of the diachronic approach. Bybee (2009, p. 17) argues that Chomsky links the theory of language universals to child language acquisition, as he considers linguistic universals to be a part of Universal Grammar, to be innate. She suggests language change as a crucial link between child language and universals (ibid). On the other hand, she adds that the "actual link between child language and the properties of languages across the world has not been established" (2009, p. 22). Similarly, Kiparsky (2008, p. 23) states that if "language change is constrained by grammatical structure, then synchronic assumptions have diachronic consequences." In analogy to Bybee he suggests explaining language change by language acquisition. Hurford (2009, p. 41) doubts the total innateness of language. The diversity of languages is caused by what is learned. He sees languages as products of a historical spiral "involving both acquisition and production, learning and speaking, and occasionally innovating over many generations (2009:51). Harris (2008:55) argues that languages are "bundles of historical accidents". Apparently, Chomsky's rejection of diachrony is unacceptable to all those who refuse to ignore the history of linguistic phenomena. Bybee (2009, p. 109), for example, points out that "all linguistic phenomena have histories which determines their present conventionalized state". The only true universals of language in the sense that they operate in all languages at all times are factors that produce cross-linguistic patterns. At the basic level of language universals there are universal mechanisms of change that create 'universal paths of change" (2009, p. 191) and even though these mechanisms work in different languages and circumstances, they are universal and they create paths of change which are often similar cross-linguistically. In this way she not only stresses the importance of diachrony but she also tries to explain the existence of universals. Blevins (2008) employs an economic motivation to justify the relevance of diachrony. Economical patterns are created by speakers in language use and once established they result in language change. According to her, there are at least three different diachronic paths through which economical patterns arise: differential phonological reduction, differential expansion of a new construction and selective analogical change (2008, p. 215).

Language universals and linguistic typology

To identify language universals, some tools must be specified. Greenberg argues that typology is a “virtually indispensable tool in the search for cross-linguistic regularities” (1963: Introduction). One of the results of the Dobbs Ferry conference was the realization “that typological classification finds its sought-for justification in the investigation of universals” (ibid). Saporta (1963) understands language universals and typology as different sides of the same coin. “We view a typology as a device which yields predictions of the type: given two languages A and B which are assigned to one type by virtue of sharing a certain feature X, and given the occurrence of another feature Y in A, we predict its occurrence in B.” (Saporta, 1963, p. 48). It is given by their nature that both language universals and language typology involve comparison of languages.

It goes without saying that language universals and typology of languages are intimately connected. They form a symbiosis, they complement each other. However, they cannot be interchanged as each of them has its specific goals and methodology of study. As Moravcsik (2007, p. 29) maintains, “while language universals are only about shared features of languages, typological statements are both about differences and similarities, the latter holding for sub-universes within the universe of all languages”. She continues: “language typology tells us what kinds of languages exist and what is possible in languages (2007, p. 35) ... in addition to it universals say what is probable and necessary” (2007, p. 36). The basic aim of typology is to classify languages into types based on selected language features. Study of language universals focuses on the features that are common for languages and identifies the scale, range, variation in which the features are set.

Croft (2003, p. 4) points out that the belief in existence of linguistic properties holding for all languages, existing beyond essential definitional properties of language and reflected by language universals is of “considerable modern currency”. The new era in universals studies started with Chomsky and Greenberg.

The generative approach to language is clearly expressed by Chomsky (1991, p. 26): “Surely if some Martian creature, endowed with our capacities for scientific inquiry and theory construction but knowing nothing of humans, were to observe what happens to a child in a particular language community, its initial hypothesis would be that the language of that community is built-in, as a genetically-determined property, in essentials And it seems that this initial hypothesis may be very close to true”. Chomsky and his followers believe that language is innate and as it was stressed in the chapter on Language Universals, they do not find comparison of languages necessary since Universal Grammar can be revealed by studying one language only – preferably English. On the other hand, since Saussure it has also been held that language is a social construct. Even though the cultural background of Victoria Beckham and an Australian aborigine are different, as soon as they share some common experience, it will be reflected in their language in a similar way. While Chomsky considers language

to be genetically conditioned, for Saussure and his followers it is society that determines language existence and development. Chomsky's formal grammar, with its interest in abstractions stands in opposition to Greenberg's empirical approach to identification of language universals. The following chart compares Chomsky and Greenberg in their approach to typology.

	Chomsky	Greenberg
background	- innate internal linguistic abilities and constraints (opposite to behaviourism) - poverty of stimulus	- anthropological relativist view of language - systematic sampling of a substantial number of languages
approach	- rationalist approach	- empiricist approach to language
method	deductive	inductive
abstraction over the data	abstracts data within languages	abstract data across languages
explanations	the biological basis is found in genetics – innate linguistic knowledge	explanations found in more general cognitive, social, perceptual etc. abilities; indirect biological bases – found in evolutionary theory

Table 29 Chomsky vs Greenberg and their approach to language universals

In Greenberg's approach, language universals result from typology. By contrast, Chomsky's starting point is a Universal Grammar in which universals are embedded. From an extreme point of view, Chomsky does not need language typology for the study of language; for Greenberg comparison and classification of languages based on empiric data is inevitable.

Tasks

- 1. Find the Universals Archive on the Internet and find the Brief introduction for Prospective Users. What is this database aimed at?**
- 2. In the same database, search for rara. In the text above you can find a table summarizing the number of language universals in individual domains. Prepare a similar table for rara and compare them. Find the definition of rarum.**
- 3. How is the Universals Archive updated with new universals? Read the following text (Plank & Mayer, THE UNIVERSALS ARCHIVE goes do-it-yourself, 2006):**

From now on, as of October 2006, the preferred mode of adding to The Universals Archive, and equally of cleaning it up, is that of Do-It-Yourself. Upon registration (you only need to send an e-mail message to thomas.mayer@uni-konstanz.de to get your personal password), on-line users will be able to feed in their own universals or any others they are aware of, as well as to add counterexamples and comments to existing archive entries. All contributions will be credited by name and email address. It is only in order to prevent too much obvious junk getting into The Universals Archive that the local management retains the right of censorship.

Do pay The Universals Archive a visit soon, to acquaint yourselves with its new DIY design; and don't forget to bookmark its new address:

<http://typo.uni-konstanz.de/archive/>

We encourage the typological community, not only to continue to consult and to work with The Universals Archive, but to make sure it is as comprehensive and as reliable as this sort of research tool ought to be. In future you will have only yourselves to blame if your own Laws of Language go unobserved and false laws continue to be paid obeisance to. We would also like to remind the readers of LT of the rubric The Universals Register, not recently the liveliest feature of this journal. What is solicited under this rubric are assessments of individual universals or preferably thematic sets – assessments surpassing in epic sweep, analytic depth, and anticipated permanency the entries and commentary mailed to The Universals Archive on the spur of the dramatic or lyrical moment.

4. Comment on the picture below.

The screenshot shows the 'The Universals Archive' website interface. At the top, there are navigation buttons for 'Home', 'Browse', 'Search', 'Rara', and 'Login'. Below the navigation bar, a search result is displayed for 'Result 2 of 157'. The main content area shows a table with the following details for universal number 11:

Number	11
Original	Consistent OV languages tend to be agglutinative in their morphology.
Standardized	IF basic order is consistently OV, THEN the morphology tends to be agglutinative.
Formula	OV ⇒ agglutinative
Keywords	order, OV, agglutination
Domain	morphology, syntax
Type	implication
Status	achronic
Quality	statistical
Basis	languages in Lehmann 1973
Source	W.P.Lehmann 1973 : 47
Counterexamples	By Frans Plank 03.08.2006, 09:49 Akkadian (E. Semitic, Afro-Asiatic): basic constituent order OV, but language is highly (intro-)flexive/fusional (e.g. <i>izzakar</i> < * <i>iztakar</i> 'he called' consisting of three morphemes, <i>i</i> - [3sg.m.], <i>z_k_r</i> 'call', <i>_ta_a</i> [Perfekt]). Akkadian word order differs from that of other Semitic languages and is generally thought to be influenced by Sumerian. (F. Kammerzell, p.c.)
Comments	By Frans Plank 03.08.2006, 09:49 See also . Should probably be the other way round: IF agglutination, THEN verb-final.

Discuss the presented language universal in as many details as possible. Follow the instructions below:

- Work with the Universal Archive and select a universal there. You can use various browsing methods to do so (e.g. number of the universal, its domain or keywords).
- Describe the universal, give its type and explain it. Use the sources given in the Universals Archive (if available). Get familiar with the terminology and with the meaning of the universal. Try to think about it in terms of the Slovak and the English languages (or any other language you speak). Give examples (and/or counterexamples) from these languages.
- Go to the WALS database and identify any chapter related to the universal. Be careful, the terminology might differ sometimes. Read the chapters with comprehension to understand the feature analysed. Complete the theoretical description of your universal.
- If needed, combine 2 or 3 features to support or reject the claim in your universal. Describe the result.

- v. It may happen that you will have to use other sources, e.g. for universals using the morphological classification of languages (e.g. agglutinative). In such a case, browse the internet and describe agglutinative languages. At the same time, compile their list. Now go back to the WALS and analyse the languages listed in the respective chapter in WALS. Describe your analysis and its results.

Summary

What is are language universals?

Discuss various types of language universals.

Compare Greenberg's and Chomsky's approach to language universals.

Describe the Universals Archive.

What are rara?

How to approach language typology?

What linguistic schools do you know?

What is contrastive and what is comparative linguistics?

What linguistic works have you read so far? What were they about? What was their theoretical background?

What is your linguistic background?

The main aim of this textbook is to get students familiar with language typology. It is impossible to reach this goal without mentioning various approaches to the field from the very beginning. And so, already in the introductory chapter, the difference between two basic approaches – the holistic and the partial ones was discussed. In addition, Haspelmath's comparative concepts were compared to Dixon's basic linguistic theory. The second chapter introduced a few predecessors of present-day language typology, including Gabelentz who gave the name to the then new field of linguistic study. The holistic approach was mentioned once again, and reference was made to the inspiring ideas of linguistic characterology (Prague linguistic school) and of Edward Sapir. A short historical overview was also included in the chapter on morphological typology. Furthermore, a new approach to lexical typology, the *Natural Semantic Method* was discussed in chapter 7. The chapter on language universals was built on comparison of Chomskyan and Greenbergian conceptions. In each chapter (including the tasks sections), the work and ideas of influential typologists are referred to. This chapter attempts to summarize the diverse approaches to language typology mentioned previously and to complete them with new information.

Prague Linguistic School

When discussing Gabelentz and his forerunners it was mentioned that they approached language as a reflection of the *Geist* (spirit) of a particular nation. They believed that each language could be classified as a whole and this approach to typology is generally labelled the holistic approach. However, the subsequent development of the field revealed that any language can be pigeonholed in various ways because typological characteristics are mixed. The interest in holistic classification of languages into language types on the basis of shared structural features was shifted to classification of selected structural features across languages. This effort engendered the period of partial typology. Thus, while holistic typology approached languages as a homogeneous whole, the starting point of partial typology is a selected structural feature, for example, word order or vowel inventory. This feature is used as a

criterion for the classification of languages. By implication, one and the same language can belong to a number of different feature-grounded classifications.

This change can, of course, be observed in Sapir's famous work *Language* (1921:127), even if he refers there to a basic plan, a certain cut, of each language, the structural genius of the language that is something much more fundamental, much more pervasive than any single feature of the language (cf. also Chapter 5, Task 6). A similar approach to language typology can be observed in Sapir's contemporary **Vilém Mathesius** who, however, lived on a different continent and was a founder of the *Prague Linguistic School*. The Prague Linguistic School had (and still has) a very strong influence on the development of linguistics not only in Central Europe but all over the world, including the already discussed Chomskyan and Greenbergian approach to language universals. Vilém Mathesius, inspired by the tradition of Humboldt, Gabelentz and other 19th century forerunners, developed a theory of **language characterology**. The main goal of linguistic characterology is to identify characteristics of a language or a group of genetically related languages using a comparative method. This is how Mathesius' follower **Josef Vachek** described it:

"... language is a system of communicative means which is implemented in concrete utterances by articulated sounds and sound attributes, or in written utterances implemented by written letters and letter-attributes. Any such utterance reacts to a certain time of extralinguistic reality, to 'things' in the broadest sense of the word. Linguistic analysis which starts on the basis of extralinguistic reality and examines the means by which the language user can express this reality, and so satisfy his communicative needs, is particularly fitted for the typological features of the language examined which make up its linguistic characterology ..." (1990, pp. 16-17)

Furthermore, Vachek points out that two fundamental areas of research in the functional-structural approach to the description of a language, that is functional onomatology and functional syntax, are best tackled if the language in question is compared to a genetically unrelated language, because it is in this way that the most striking features of a language can be highlighted.

Another representative of the Prague School who devoted his attention to language typology was **Vladimír Skalička**. His approach was both deductive and diachronic. He applied language typology to individual languages but also to language families. He was deeply interested, among others, in Finno-Ugric, Turkic, Bantu, Balkan and Germanic languages. A typical feature of his approach is the so-called '**construct**' terminology in which construct or model or ideal extreme has only a methodological value. Any of the existing languages can be classified as a construct. Ideal constructs are made of a number of mutually supporting features. There are five ideal constructs: the agglutinating and the inflective constructs are identical to the classical typology proposed by Sapir. The isolating construct corresponds with Sapir's isolating/analytic type. The introflective and

polysynthetic types are Skalička's innovation. This is how the properties of individual constructs can be described (based on (Čermák, 2011, p. 230)):

Language construct: agglutination

- grammatical meanings are expressed by a chain of affixes
- transparent and diagrammatic morphology
- affixes express one grammatical category
- affixes are phonologically independent, they have consonants
- no allomorphy
- no suppletivism
- no classification into word classes
- no category of gender/class
- no grammatical synonymy
- no grammatical homonymy
- specific derivational morphemes
- no agreement in nominal constructions
- nominal clauses
- no lexeme of possessive predication (verb 'to have')
- fixed word order

Language construct: inflection

- grammatical meanings expressed by one affix
- affixes express more grammatical categories, cumulation
- affixes are phonologically dependent, no consonants
- allomorphy
- suppletivism
- classification of words into word classes
- existence of the gender/class category
- grammatical synonymy
- grammatical homonymy
- stem inflection
- flexive morphology applied in word formation
- verbal clauses
- free word order

Language construct: introflexion

- grammatical meanings expressed by infixation or non-concatenative modification of the lexical morpheme
- + some features of inflection: word classes, inflectional morphology, agreement, flective morphology, verbal clauses, free word order
- isolative construct
- synsemantic words express grammatical meanings
- monosyllabic words
- lexical homonymy and polysemy
- no word classes
- no morphological word formation
- finite clauses
- fixed word order

Language construct: polysynthetic

- grammatical meanings are expressed by combination of autosemantic lexemes
- no synsemantic words
- no morphology
- no word classes
- compounding
- compound clauses
- fixed word order

Language construct: isolating

- monosyllabic words
- absence of affixes
- conversion
- numerous function words
- free word order

Peter Sgall developed and extended Skalička's approach. In his view, each language type has one dominant feature from which other features can be deduced – certain characteristics of language are inherently linked. While the structure of a language is made up of various language type characteristics, one type can be dominant. As Sgall suggests, "the predominance of a particular type in the structure of a language may be interpreted in terms of linguistic economy: adherence to a

single type (either grammar words, or affixes, or alternation, or order) as the means of encoding relational meaning would appear less costly than an unprincipled deployment of several types side by side.” (Shibatani & Bynon, 1995, p. 8). Sgall employs Skalička’s notion of construct in order to focus his attention on the definition of language type. He defines language type as a collection of linguistic properties characterized as follows (Sgall P. , 1995, p. 50):

- a) the properties are intrinsically connected by probabilistic implications of the form ‘if a language has the property A, then it probably also has the property B’;
- b) the types are ideal extremes (constructs) not fully attainable by existing languages; the latter come more or less close to one (or more) of the ideal types;
- c) properties of different types are combined within the structure of every existing language.

The Prague Linguistic School is a school of structural and functional linguistics in which the language system is conceived of as a system in its own right. Thus, the typology developed within its framework is structural and functional. This diachronic approach can be found, for example, in a number of papers by Skalička.

Leningrad Typology Group

Another typological group which grew in the linguistic traditions of Europe is the *Leningrad (St Petersburg) Typology Group*. It was founded by **Alexander Xolodovič** from the Institute of Linguistics of the Soviet Academy of Sciences. The work of this group was typical of the so-called **collective** method, which basically means that there was a team of linguists with a leader, coordinator. The leader decided on the topic which was further discussed within the research team. The discussion resulted in a questionnaire on the basis of which individual specialists described the language of their respective expertise. The result of this kind of work was usually an anthology of papers, e.g. *Tipologija kauzativnyx konstrukcij: Morfoložičeskij kauzativ* (Typology of Causative Constructions: Morphological Causative (Xolodovič 1969)).²¹

Two members of the Group, Nedjalkov and Litvinov (1995, pp. 237-238), identify seven stages in the collective method: discovering the problem; drafts on the ontology of the phenomenon chosen for investigation; drafting papers on the problem in individual languages; formulating the problem; conference on the theme; rewriting the chapters on the particular languages in accordance with the questionnaire; reviewing the submitted papers; and editing a collective monograph.

Certainly, members of the research group were not specialists without their own ideas. It happened that a language specialist did not accept the program without questions, which meant that meetings

²¹ For details cf. (Song, 2001, pp. 340-341).

of the research group were often endless, tense and dramatic (Nedjalkov & Litvinov, 1995, p. 218). With Xolodovič managing the meetings they would take the form of extreme brainstorming. As Nedjalkov and Litvinov claim (1995, p. 218) “the participants did not subordinate themselves to a personality, however authoritative, but rather to a programme that had grown out of ideological conflicts, and which the participants now regarded as their own programme rather than as one imposed upon them from outside.” Xolodovič as a leader co-authored all the manuscripts that were published as a result of the team work and it often happened that he rewrote some parts to make the manuscript match the scheme. The basic aim was to introduce “new knowledge within a compact and coherent framework” (ibid). The organization of the team was dual – a leader who coordinates and distributes the tasks and members of the group who work on the allocated topic or problem.

In the course of time the group was replaced and swallowed by the *Section for the Theory of Grammar and Typological Research* headed by A.V. Bondarko. After 1992, however, the work of the group was resumed. It works as an informal group nowadays. The major underlying assumptions of the group with respect to typology were formulated by Nedjalkov and Litvinov as follows (1995, p. 260):

1. Typology does not treat natural classes, nor does it postulate any ‘natural types’. Therefore, it does not have to be global. It is a typology of phenomena, or categories, or sub-systems. What is known as ‘sub-system typology’ is in fact typology proper.
2. Typological description of language sub-systems is based on a cluster of features; a description obtained is such that it can be reformulated in terms of another cluster of features. It is theoretical in the sense that it admits, in principle, of varying interpretations.
3. In order to be typological, a theory must reveal and lay emphasis on the diversity of all language material as relevant.

The list of publications of the Leningrad Typology Group is very long, and recently the complex methodological approach of the Group has been applied also in individual research, e.g. by Geniušienė who studied the typology of reflexives.²²

The UNITYP research group

The **UNITYP** (Language Universals Research and Language Typology) research group started in 1972 at the university of Cologne, Germany. It was founded by **Hans Jacob Seiler**. Supported by the *German Research Council Fund* the group also cooperated with the University of Guadalajara in Mexico and the University of Geneva in Switzerland. Members of the research group met regularly and their meetings

²² For list of publications and further information about the Leningrad Typology Group cf. Shibatani and Bynon (1995).

were recorded. These minutes were used to identify the directions in the research. As Seiler points out (1995, p. 274), the carefully edited minutes of the project meetings (in typescript) are the characteristic feature of the UNITYP's activities. There are four series within which most of the work has been published. They cover eight cognitive-conceptual dimensions (ibid): NOMINATION, CONCOMITANCE, DETERMINATION, POSSESSION, APPREHENSION, PARTICIPATION, SITUATION, LOCALIZATION. Seiler defines research into language universals as a task that consists in "uncovering the ways in which cognitive conceptuality of a universal status is constructed and in demonstrating the ways in which such construal is represented in individual languages" (Seiler, 2000, p. 11). The key term of his approach is **functional denominator** that is understood as an umbrella term for a number of structural phenomena in any given language, although differing in form and/or meaning. For example, a common function of relative clauses, genitive constructions, quantifiers and demonstratives in Ancient Greek is to represent identification of an object or item. He orders them according to similarities and differences, e.g. adjective and genitive have more in common than e.g. adjective and relative clauses. It results in a continuum from relative clauses to demonstratives. (ibid). The construct of continua enables one to compare data from various languages within the frame of the common denominator of cognitive-conceptual domain. At the same time, Seiler views language as a problem-solving system. He distinguishes between semasiology (language-specific meanings) and onomasiology (cognitive-conceptual content) (Seiler, 1995, p. 321). This makes his approach more comprehensive. As Seiler puts it, one approach is language specific, observational and comparative while the second rational and conceptual. Language specific data are visible at first sight. On the other hand, concepts are not directly observed but "...are first hypothetically posited and then tested for their utility in understanding language and languages" (Seiler, 2000, p. 23).

The EUROTYP project

EUROTYP stands for "Typology of Languages in Europe". This project on language typology was initiated by four typologists: Johannes Bechert, Claude Buridant, Martin Harris and Paolo Ramat. This large-scale project of the *European Science Foundation* was aimed at the study of "the patterns and limits of variation in nine focal areas: pragmatic organization discourse, constituent order, subordination and complementation, adverbial constructions, tense and aspect, noun phrase structure, clitics and word prosodic systems in the languages of Europe" (Dahl, 2000, p. v). One of the ambitions of the project was to identify *Sprachbünde* within Europe. The project involved more than 100 linguists from more than 20 European countries and the USA. The research outcomes were published in nine volumes corresponding with the above-mentioned foci. Each volume has its own

editor (G. Bernini, A. Siewierska, N. Vincent, J. Feuillet, J. van der Auwera, O. Dahl, F. Plank, H. van Riemsdijk, H. van der Hulst). Moreover, the project initiated new interesting typological studies.

One of the basic research tools of the project were questionnaires whose description is still available at the website of the Department of Linguistics at the *Max Planck Institute of Evolutionary Anthropology*. Even though the department does not exist any more, its former members are still very active in typological studies. Among their interests belong, for example, diversity of human languages, language documentation and description of languages. They have also developed various tools which are used by typologists all over the world, e.g. *WALS* (World Atlas of Language Structures and *APICS* (Atlas of Pidgin and Creole Language Structures). One of the results of the *EUROTYP* project was identification of the so-called *Charlemagne area* (cf. chapter 7).

Canonical typology

Canonical typology has been developed at the *University of Surrey* by members of the *Surrey Morphology Group*. The main objective of research into canonical typology are phenomena which vary across and within languages. The main methodological tool is the use of scales (called **criteria**) which “relate to the presence or absence of a property or characteristic, or the degree to which a particular property is manifested by a particular set of data” (Group, 2016). Canon means a general law, rule, principle, or criterion by which something is judged. Canonical instances are those which most closely match the canon, the rule, the principle, they are the best examples. Criteria are structured in such a way that when a dataset meets one of the criteria it is more canonical along that parameter than those datasets that do not (ibid). Canonical typology establishes the logical beginning and end of such scales. By implication, the non-canonical properties are identified, too. When a phenomenon is studied in canonical typology, criteria (the above-mentioned scales) associated with that phenomenon are set. If an instance of that phenomenon is canonical, it means that it meets all the criteria and it is clear and indisputable instance of the phenomenon under investigation. As Corbett (2007, p. 9) points out, canonical instances are unlikely to be frequent. He also mentions (ibid) his personal communication with Joanna Nichols who puts it in the following way: Canonical constructions are all alike; each noncanonical construction is noncanonical in its own way. This is compared to a sentence from *Anna Karenina* by Tolstoy: ‘All happy families are alike; each unhappy family is unhappy in its own way’. There are three key concepts in canonical typology: the already mentioned **criteria**, the **canonical base** and the **canonical ideal**. “The base defines the broad space of particular linguistic phenomenon to be described by the typologist” (Brown & Chumakina, 2013, p. 3).

Linguistic objects are understood as entities in a space. They differ from each other in their proximity to a point of convergence. The point of convergence represents the canonical ideal. The space in which linguistic objects, phenomena are placed can be defined in two ways: either by long theoretical

discussion or by a definition of the notional starting point. This notional starting point is the base. For example, a notional starting point for the investigation of inflectional class can be expressed by a number of the following key properties (Brown & Chumakina, 2013, p. 4):

“For a given language an inflectional class is one of a set of classes which cross-cut syntax and which define the forms for words belonging to the same syntactic category.” It is obvious that the base is usually a very broad definition of the phenomenon under investigation, e.g. in our example we learn that inflectional class is something that relates word forms and syntactic categories. The domain of investigation – inflectional class is defined by the base. The dimensions of the domain are defined by criteria. The main task of the criteria is to state “which points represent the more canonical, and less canonical instances along the dimension” (ibid). The point of convergence in the dimension set by criteria is the canonical ideal.

When discussing Canonical Typology, the so-called **correspondence** problem should be mentioned, too. It refers to the situation which has been mentioned several times in this textbook – the correspondence between similarly named features in different languages (cf. Haspelmath’s comparative concepts). As the representatives of this approach claim, the canonical typology “seeks to avoid the tendency to use linguistic terms with vague and shifting definitions by placing emphasis on the criteria used to associate particular linguistic phenomena with cross-linguistic categories”. (Brown & Chumakina, 2013, p. 3).

Besides the canonical approach to language typology, the Surrey Morphological Group has also developed a set of databases that can be used for typological studies, e.g. the *Oto-Manguenan Inflectional Class Database*, the *Surrey Morphological Complexity Database*, the *Saanich Verb Database*, etc.

Košice typology group

The **Košice typology group** is based at the Department of British and American studies, P.J. Šafárik University. Since 2005 every three years a conference is organized by the group members. The title of the conference series, Typology and Universals in Word-Formation, illustrates the interests of the researchers. The group was established by P. Štekauer and his students and colleagues. Their work is strongly influenced by the ideas of the Prague Linguistic School, especially by an innovative approach of M. Dokulil to word-formation. This approach is called onomasiological. Some time ago a new, so-called cognitive approach started in linguistics and its principles are very close to Dokulil’s onomasiology. One of the first achievements of the group was the publication of the monograph *Word-formation in the world's languages. A typological survey* by Štekauer, Valera and Körtvélyessy (2012), the first typological work on word-formation, covering 55 languages (cf. also Chapter 5 in this

textbook). This work was followed by a typological analysis of phonetic iconicity and evaluative morphology by L. Körtvélyessy (2015). The cross-linguistic exploration in this monograph is based on data of 200 languages of the world. In addition, this monograph also pursues the delimitation of Standard Average European and compares its results to those obtained by evaluating other morphosyntactic categories within the EURO-TYP projects and some other monographs, e.g. Haspelmath (1998, 2001). Languages are compared on the basis of a new parameter, the so-called **saturation value**, that – generally speaking - reflects the structural richness of a system examined. Members of the team have also co-edited extensive volumes on compounding, affixation and evaluative morphology that contain a large number of chapters describing these phenomena in various languages/language families.²³ At present, the group members have been working on a large-scale typological project aimed at research into the languages of Europe based on the evaluation of 120 word-formation features and relying on the use of the parameter of saturation value.

Considerable attention has been paid to the phenomenon of phonetic iconicity. An analysis of 50 languages of the world (Gregová, Körtvélyessy, & Zimmermann, 2010), resulted in the rejection of Universal #1926 as presented in the Universals Archive. The authors demonstrate that phonetic iconicity in word-formation is of areal rather than of universal nature. Körtvélyessy (2015) undertook an experiment that covered five age groups of respondents representing three different languages. The experiment proved the tendency of language users to be more ‘iconic’ at lower age and less ‘iconic’ at higher age. The experimental results suggest that informants of a higher age prefer to apply linguistic knowledge instead of feelings.

Panocová (2015, 2016) compares neoclassical formations in typologically distinct languages from the perspective of onomasiological theory of word formation.

Ongoing typological research at the level of syntax aims to test formal means that speakers of selected European languages employ to activate the selected cognitive chains (Non-Agent Argument chains, in particular), to identify case frames (cognitive-coding types or CC types) realized by those formal means and to find if/how CC types coincide with the genetic and morpho-syntactic typological classes of languages (Janigová).

Tasks

1. **Read the following paragraph by Sgall. Compare it to Haspelmath’s comparative concepts. Where does Sgall see the future of language typology? Have his words come true?**

²³ R. Lieber and P. Štekauer (2009) *The Oxford Handbook of Compounding*; R. Lieber and P. Štekauer (2014) *The Oxford Handbook of Derivation*; N. Grandi and L. Körtvélyessy (2015) *The Edinburgh Handbook of Evaluative Morphology*.

Let us add one more remark to the deductive, or formal, approach. It has, of course, its limitations; apparently, these are connected with the fact that undefined primitive notions (regarded as sufficiently clear) are necessary here. Usually such notions as "lexical unit", "word", "affix", "base" belong to them. But, in the empirical domain, these notions are not always very clear; in some cases, if we want to find intrinsic distinctions (and not to accept solutions imposed by orthographical conventions, grammatical tradition, etc.), we must use criteria connected with the presence of different types within a single language.

.... But even with these limitations, the deductive approach certainly can and will be useful for the further development of typological studies, in connection with the algebraic theory of descriptions of language as well as in connection with the study of the development of languages (since typological properties are one the main factors of development) (Sgall, 2006, p. 207).

2. Compare Skalička's idea of an ideal construct to Chomsky's ideal speaker/listener. What is the biggest problem with ideal phenomena?
3. Compare morphological classifications proposed by Sapir and Skalička.
4. This is how Nedjalkov and Litvinov (1995, p. 235) illustrate their work within the collective method. Comment on the picture.

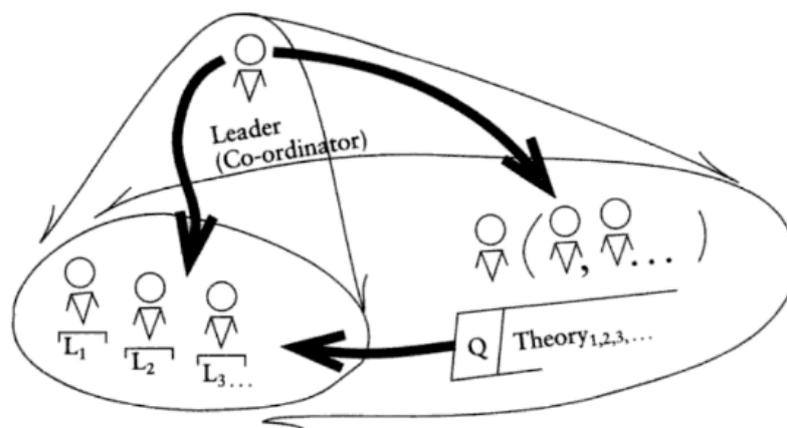


Figure 24 The collective method in the Leningrad Group

5. The following text comes from a review of Seiler's monograph *Language Universals Research: a Synthesis*. The review was written by Torsten Leuschner and published in *LoLa*. Read it and summarize your knowledge of the work of H.Seiler.

According to Seiler, "universals research involves nothing less than a quest for the essence of language" (p.11, 194). He therefore regards it as by nature provisional and open-ended, and insists that readers should not expect him to provide a final (and finite) catalogue of language universals that could somehow stand as a fixed body of knowledge independent of any particular theoretical framework (p.11). Rather, Seiler's book is an account and a synthesis specifically of the notions, methods and results of UNITYP. After the Introduction (chapter 1) and a short survey of the principles and working methods of the Project for

rapid orientation (chapter 2), Seiler compares the UNITYP framework with other, related approaches to universals and typology, in particular 'Greenbergian' functional typology (chapter 3), and gives a short overview of the history and publications of the project, including a list of the many, often distinguished linguists who have been its members, associates or guests (chapter 4). UNITYP's methods and results are then exemplified in two chapters about the 'dimension' of IDENTIFICATION (dimensions being given in capital letters in UNITYP usage). The first (chapter 5) approaches IDENTIFICATION from the semasiological point of view, i.e. it reviews and inductively seeks regularities among structures from several languages (mainly German and Ancient Greek, but also Tolai, Samoan and Cahuilla) that are said to fulfill an 'identifying' function such as demonstratives, articles, relative and attributives constructions, quantifiers, numerals, pronouns etc.; this chapter takes up more than 100 pages, almost half of the entire book. A shorter section follows in which IDENTIFICATION is approached from an onomasiological point of view, i.e. by abductive reflection rather than sampling of data (chapter 6). Other dimensions that figured in UNITYP research (often involving several additional languages) are then reviewed briefly, viz. APPREHENSION, NOMINATION, NUMERATION, POSSESSION, LOCALIZATION, PARTICIPATION, and OPPOSITION (chapter 7). The book is summarized in seventeen short sections on the terminology, basic notions and ideas of UNITYP (chapter 8), and closed by a list of references, indices of topics and of languages and language families, and a list of diagrams and tables.

What distinguishes Seiler's book from earlier publications in the same series is, of course, its focus on the general methods and principles of UNITYP. Seiler brings an ingenuous approach to the job by rejecting the easy option of showcasing previously published research (except at a late stage, viz. chapter 7) and opting instead for a fresh case-study of IDENTIFICATION. Placing the case-study central among more abstract discussions of alternative approaches, terminology, the history of the project etc., ensures a high level of theoretical and methodological reflection throughout the book. As Seiler himself recognizes, this special manner of presentation is necessary not least because research in UNITYP differs quite substantially from most other language universals research (p.11). Its main distinctive trait, according to Seiler, is its teleonomic view of language, i.e. "its conception of language as a directed activity, as having 'purposive function'" (p.17); the two main purposes that language is seen as fulfilling are "construction of content" and "representation of content" (ibd.). At the risk of oversimplifying and de-nuancing Seiler's own presentation, 'dimensions' like IDENTIFICATION, POSSESSION etc. can be described as the conceptual building-blocks for the cognitive construction of the content that individual languages need to represent by means of linguistic forms, constructions etc. While acknowledging that "(p)urposive function is not a notion particularly favored by current mainstream linguistics", Seiler can nonetheless cite Jakobson's means-ends model of language and Benveniste's view of "the spoken word [as] an output of mental operations" in support of his views (ibd.). Another researcher in the European structuralist tradition to whom UNITYP looks for inspiration is Eugenio Coseriu: quoting at length Coseriu's distinction between possible, essential and empirical universals (p.25f.), Seiler emphasizes that universals cannot be discovered by empirical observation alone (p.26). The latter point is an allusion to the 'implicational universals' as proposed by functional typologists like Joseph Greenberg. While Seiler regards implicational universals as "suggestive generalizations" (p.26), he accuses proponents of this approach (above all Greenberg himself and Bernard Comrie) of "lack of theoretical rigour" with regard to the concept of universality in their research (p.27). However, Seiler fails to take into account such foundational works as Comrie (1989) and Croft (1990), and the gulf between their approach and UNITYP might have seemed less unbridgable had Seiler pointed out that 'Greenbergian' researchers often situate their work in the tradition of Wilhelm von

Humboldt (e.g. Hopper / Traugott 1993: 18f.). That Humboldt and in particular his concept of *energeia* are not mentioned at all in Seiler's present book is surprising indeed, since Seiler himself stresses UNITYP's indebtedness to Humboldt and *energeia* in another, earlier presentation (1995: 299, 320), and so does Lehmann in his study of relative clauses (1984: 20f.).

6. Go to the following website: <http://www.eva.mpg.de/linguistics/past-research-resources.html?Fsize=0%2C%20%40> and find out what kind of typological surveys and databases were developed at the former department of linguistics of Max Planck Institute. How are they connected to the EUROTyp project?
7. This is how Corbett describes canonical typology. Read the paragraph and summarize your knowledge of the subject.

Canonical typology (Corbett G. G., 2013, p. 48)

Adopting a canonical approach means that we take definitions to their logical end point, and this enables us to build theoretical spaces of possibilities. Only then do we investigate how this space is populated with real instances. The canonical instances are simply those that match the canon: they are the best, the clearest, the indisputable ones. Given that they have to match up to a logically determined standard, they are unlikely to be frequent. They are likely to be rare, and may even be non-existent. This is not an issue. The convergence of criteria fixes a canonical point from which the phenomena actually found can be calibrated. We may then go on to an investigation of the distribution of canonical and less canonical phenomena in terms of their frequency. The canonical approach has been applied to phenomena in syntax, notably agreement (Evans 2003, Corbett 2003, 2006, Comrie 2003, Polinsky 2003, Seifart 2005, Suthar 2006) and to phenomena in morphology, by Baerman, Brown & Corbett (2005: 27-35), Spencer (2005), Stump (2005, 2006), Corbett (2007), Nikolaeva & Spencer (2008), Stump & Finkel (2008) and Thornton (2008).² Within this approach, I have proposed (Corbett forthcoming) a specific set of converging criteria for canonical features and values, concentrating on the genuine morphosyntactic features. Various non-canonical behaviours have been identified. To take one example, we find that a lack of robust formal marking underlies non-autonomous case values. Rather than being restricted to one feature, like case, parallel behaviour recurs in different features, sometimes with different labels.

Summary

What is holistic and what is partial approach to typology?

Discuss Sapir and his typological approach.

Compare Greenberg's and Chomsky's approach.

Discuss the Prague Linguistic School.

Discuss the Unityp and Eurotyp projects.

How do typologists work?

What methods of linguistic research do you know?

What is introspection?

What linguistic databases do you know?

Describe research in any linguistic field.

Research strategies of typological research

Each linguistic research starts with identification of a phenomenon to be analysed. The phenomenon can be studied by means of various theoretical approaches. The main aim of this stage is the delineation of the phenomenon in question. If the research is typological, the subsequent step to be taken is the identification of a language sample. By implication, typological research should follow two directions:

1. Theoretical – the main aim is to identify the phenomenon to be analysed
2. Language sampling and data collection – the main aim is to collect reliable data from a balanced sample of languages.

Each direction faces its problems. Let us take up first the theoretical stage. According to Croft (2003, p. 4), the standard research strategy for typological research includes the following steps:

1. determine the particular semantic-pragmatic structure or situation type that one is interested in studying;
2. examine the morphosyntactic construction(s) or strategies used to encode that situation type;
3. search for dependencies between the construction(s) used for that situation and other linguistic factors: other structural features, other external functions expressed by the construction in question, or both.

The result of this strategy should be a detailed description of the phenomenon under investigation. Generally, this strategy is in accordance with the fact that the already mentioned *tertium comparationis* has to be specified. Its identification raises a serious methodological problem that has been frequently mentioned in this textbook and to which Song (2001, p. 14) refers as the “**peril of typology**”. You may also remember Corbett’s **correspondence problem** from the previous chapter which leads us to Haspelmath’s **comparative concepts**, Dixon’s **descriptive categories** and Sgall’s **undefined primitive notions**. All these notions reflect one and the same fact: the compared phenomenon should be expressed in terms that are identifiable in each language, i.e., it requires cross-linguistic categories. The problem is that despite various efforts, they have not been identified yet. If

we perceive language as a set of bilateral linguistic signs in the Saussurean sense, there are two possibilities. The first, usually applied by generative grammar, to make use of formal descriptive categories. The problem is that the same set of categories cannot be applied to every language as languages vary in many respects. This is not an issue for generative grammar owing to its postulate of a Universal Grammar that underlies all languages. The second option is usually preferred by the functionalist approach to language universals. Croft (2003, p. 14) has recourse to semantics, in particular, in semantics that covers “a large area including pragmatics, discourse-defined connectives etc.” (ibid). Contrary to this, Newmeyer (2007) and Song (2001) point out that the concept-based or meaning-based categories are not universal either and therefore they insist on comparison based on formal categories. A different way out of this problem is proposed by Haspelmath (2010) who suggests the application of comparative concepts as a kind of metacategory.

It should be pointed out that even though typologists apply the semantic approach, they usually follow the semasiological method – they proceed from form to meaning. Semasiology is in opposition to onomasiology which puts the emphasis on meaning, on the way cognitive categories are formally expressed. Therefore, onomasiology starts from the conceptual level and as such it is complementary with cognitive linguistics. No doubt, onomasiology and cognitive linguistics can support the identification of categories comparable in all languages. The problem, however, is that ‘mainstream’ linguistics relies on the semasiological approach. One of the reasons bears on the traditional account of languages in descriptive grammars all of which are semasiologically grounded. By implication, the collection of onomasiological data is much more laborious. This prejudice has had a very long tradition and may be found as early as the beginning of the 20th century. Bloomfield (1933), for example, stressed the importance of formal analysis of language. Meaning in his view is illusive. This kind of view influenced generations of linguists. In Chomskyan transformation and generative grammar meaning plays hardly any role, either.

The question of the methodological approach to *tertium comparationis* is also vital to answering the next question: Who should be asked for data? Where should typologists seek their data? What kind of sources should be used? These questions will be tackled below. First, however, let us pay attention to the closely related problem of obtaining a representative sample of languages.

Language sampling

Since typology examines cross-linguistic variation, it requires an appropriate language sampling procedure and data collection procedure. Both of these procedures face serious obstacles. An ideal language sample is balanced from the perspectives of genetic relationships and areal distribution. During the sampling process, various bias-related problems can occur. Bakker (2013) lists the following bias possibilities:

- bibliographic bias – caused by the fact that more than two thirds of world languages have not been described yet;
- genetic bias – means overrepresentation for language families that are well known;
- areal bias – reflects the situation when languages are part of the same linguistic area;
- typological bias – a sample contains a disproportionate representation of languages of which the type has a direct or indirect bearing on the values of the research parameter;
- cultural bias – “there is a skewing in a sample as far as the representation of the world’s cultures is concerned” (Bakker, 2013, p. 108).

In typological studies, linguists do their best to collect data on as many languages as possible. Actually, the size of data is one of the facts which make typology different from contrastive linguistics. Individual chapters in the WALS database, for example, typically cover more than 100 languages. For illustration, the chapter on consonant inventories analyses more than 500 languages. The samples can vary in their nature. Essentially, six types of language samples are used:

1. **Variety** sample – selects languages that “have evolved independently from each other for a long enough time to have developed different strategies for the grammatical expression of the phenomenon under study” (Croft (2003, p. 21)). The advantage is that “linguistic variables are explored about which not much is known in advance. In such cases, it is precisely the variation among the values for the respective variables that we want to know” (Bakker, 2013, p. 104).
2. **Probability** sample – is usually “relatively small in size between 50 and 200 languages, and will vary with what is known beforehand about the range of values for the relevant linguistic variables and their stability” (Bakker, 2013, p. 104). The sample consists of values representative for those genealogical units which “are old and historically diverse enough so as to be reasonably independent with regard to the typology under study” (Bickel, 2008, p. 221).
3. **Convenience** sample – also called opportunity samples – available data are ‘grabbed’.
4. **Quota** sample – analysed material is divided into categories. Each category must be represented by a certain number of objects.
5. **Judgment** data – “the selection depends upon investigator’s expert knowledge of the subject matter” (Bell, 1978, p. 128).
6. **Random** sample – represents independent data selection. For Bickel it is the “hallmark of statistics” (2008, p. 3) as all languages had the same a priori chance of being included into the sample.

Preferably, samples should be created of unrelated languages: no two languages are from the same family or area. This principle is very difficult to comply with. Dryer (1989) suggests the following

procedure: (i) group languages into genera; (ii) divide languages in genera into 5 continental territories, including Africa, Eurasia, Australia-New Guinea, North America, South America; (iii) count the number of genera of each of the relevant linguistic types being investigated within each continental territory; (iv) determine how many of the five areas confirm the hypothesis. His approach is further elaborated by Bickel (2008) who pays special attention to the actual distributions within genealogical units at each taxonomic level.

Data collection

Gathering information about languages in a sample usually relies on either reference books or questionnaires / data sheets, or both. Both techniques have their hiccups. Reference books may be incomplete in terms of the categories examined, or the description of the particular category may be totally absent due to the lack of expertise of its author or due to the absence of data available to the author; the data can be misunderstood and misinterpreted due to different terminologies used by various theoretical schools, or due to the lack of relevant explanatory comments by the author of a descriptive grammar; certain linguistic categories may manifest themselves differently in different languages or language types; and the limited amount of reliable data on a particular language may result in overgeneralization. All this is especially true of reference grammars covering endangered languages and produced by fieldworkers. They, for obvious reasons, completely depend on the information from native speakers who, once again, for obvious reasons, are usually not linguists. Questionnaires, on the other hand, are time-consuming on the part of an informant, and money-consuming on the part of the researcher. It is very difficult to create a reliable and comprehensive questionnaire that balances the expectations of a researcher, on one hand, and the extent of the desired questionnaire data psychologically acceptable to and manageable by an informant. The best way to avoid these traps is to combine both reference books and questionnaires.

To sum up, when it comes to data collection, the procedure usually starts with the identification of written sources on the specific language. A typical tool used here is a simple Google search. It is very helpful if the language is covered in databases like Ethnologue, WALS and Glottolog because they provide a linguist with a list of references, too. If this is not the case, one can continue in the Google search. In some cases, the alternate names of a language must be taken into consideration, e.g. the language *Dhivehi*, spoken in India, is also called *Maldivian*. In addition, there are genetically unrelated languages labelled with the same name, such as *Buli*. It is either a Niger-Congo language spoken in Ghana or an Austronesian language spoken in Indonesia. The next step is gathering the identified sources and reading them with the aim to find answers to the specific research questions. The described procedure is very time-consuming and does not guarantee that the typologist finds an answer to the research questions. Or, it may be that some additional clarification is needed. At this moment, a new Google search starts with the aim to identify the author of the written source, or some

other specialist in the particular language. Again, this can be a very long process requiring extensive mail communication because, unfortunately, the return rate is rather low. As soon as the language expert is identified a new email exchange can start in which the typologist explains the research objectives and the theoretical background. This should facilitate the language expert to supply satisfactory replies. Fortunately, this long and demanding process is supported by the existence of various databases. Some of them have been already mentioned in the textbook, as, for instance WALS, Ethnologue, Universals Archive, the databases of the Surrey Morphological Group and reduplication database.²⁴ Here are some others:

LLOW - Languages of the World – organizes and describes languages of the world and its goal is an extensive description of world languages of linguistic-genetic (language families, clusters, ...) and sociolinguistic (status as official language, minority language, ...) criteria, as well as social-, geopolitical and geographical parameters (number of speakers, geographical extension, ...).

Glottolog – aims at providing comprehensive reference information about the world's languages, especially the lesser known languages.

The Pavia Typological Database – aimed at the typological documentation of morphosyntactic phenomena in languages belonging to the Euro-Mediterranean area

APiCS - Atlas of Pidgin and Creole Language Structures

WOLD - The World Loanword Database

The last two databases were developed at the former Department of Linguistics of the Max Planck Institute. The Department was also very active in other respects. On their website one can find a collection of questionnaires used in the EUROTYP project. The website also provides guidelines to the questionnaires and stimulus kits for the data elicitation. Furthermore, the Department members developed a consistent and widely accepted standard for the interlinear glossing of text known as **Leipzig glossing rules**. This standard is used in this textbook (cf., for example, the chapters on morphological typology and syntactic typology). Another type of activity related to the previous ones is language documentation and fieldwork covering more than 60 languages.²⁵

These achievements give a new dimension to typological research. A typologist aims at comparison of a large number of languages. As has been already explained, the typologist does not (and cannot) speak all the languages under investigation. This is compensated by cooperation and consultation with

²⁴ Cf. chapter 1, task 1.

²⁵ <http://www.eva.mpg.de/linguistics/past-research-resources/documentation-and-description/languages-and-language-groups-which-we-are-studying.html?Fsize=0>

language experts. Language experts usually collect their data in the field (especially in the case of endangered languages). This demanding work is reflected in the existence of a special linguistic branch called **fieldwork**. A fieldworker produces a **language description** complemented with **language documentation**. Language documentation deals with the collection, processing and archiving of linguistic data as, for instance, texts, word lists, various recordings – both audio and visual.

Fieldwork and language documentation

The importance of fieldwork and language documentation has been growing hand in hand with the interest in language typology. A rapid growth of this interest can be observed in the last decades. This fact is also reflected in a number of descriptive grammars. In 2011 Springer published the *Handbook of Descriptive Linguistic Fieldwork* and in 2012 it was followed by *The Oxford Handbook of Linguistic Fieldwork* in which linguistic fieldwork is discussed across nearly 500 pages. The following lines focus on those facts that are crucial to language typology. First of all, descriptive linguistic fieldwork should not be confused with language typology itself. This idea is pointed out explicitly by authors of the *Handbook of Descriptive Linguistic Fieldwork*. They discuss fieldwork from a wider scope than other linguists and define fieldwork in terms of both what it is and what it is not. According to them, descriptive linguistic fieldwork is (Chelliah & De Reuse, 2011, p. 8):

1. Data collection for the purpose of the documentation and description of a language.
2. Data collection through interaction with speakers.
3. Data collection in situations where speakers are expected to use the language naturally.

Descriptive linguistic fieldwork is **not**:

1. Data collection only through introspection.
2. Data collection only through examination of written documents or written corpora.
3. Data collection only through controlled lab experiments.

It is obvious that the NOT part of this description refers rather to language typology, especially when it comes to points 1 and 2. According to the authors of the *Handbook*, linguistic fieldwork is defined as “the investigation of the structure of a language through the collection of primary language data gathered through interaction with native-speaking consultants.” (Chelliah & De Reuse, 2011, p. 7) By implication, fieldwork requires personal contact with native speakers. This idea is supported by Aikhenvald who claims that linguistic fieldwork involves “venturing into a community where the language is spoken, collecting the information, and providing a comprehensive analysis and written documentation of the language” (2009, p. 1). She stresses the importance of the first-hand knowledge of diverse languages which is essential for the understanding of how human languages work. Fieldwork is the ‘backbone’ of linguistics. On the other hand, those who make use of the data collected by

fieldworkers are labelled as 'arm-chair typologists'. (Aikhenvald, 2009, p. 2). The problem, however, is that the present-day typology requires a large number of languages to be compared (cf., for example, the chapters in WALS). If the number of languages is small (about 3-5), the linguistic analysis is carried out within the contrastive framework, not the typological. This raises the question of the number of languages that can be documented by one fieldworker. In Aikhenvald's view, being a typologist automatically means being a fieldworker. This is not in accordance with the above-mentioned approach by Chelliah and De Reuse. The boundaries between typologists and fieldworkers are fuzzy, they do not exclude each other. Importantly, being a typologist is not necessarily preconditioned by being a fieldworker.

From the historical perspective, Chelliah & De Reuse (2011, p. 33) distinguish five types of linguistic fieldwork:

- Christian missionary fieldwork: typical especially during colonisation first in Africa and Asia, later in Latin America, North America, Australia, Papua New Guinea and the Pacific. The most motivating factor was that Christianity required communication in the language of people for missionary purposes. So, for example, the first systematic observations of the spoken Indian languages were carried out by the Jesuit Thomas Stephens.
- the "gentleman-scholar" fieldwork: explorers or travellers, military or colonial personnel, ambassadors, politicians, school teachers with deep interest in human languages did fieldwork as their hobby. For instance, the English lawyer, antiquary and naturalist Daines Barrington found a few native speakers of Cornish in the town of Mousehole.
- fieldwork in less than optimal circumstances: this is fieldwork in unpleasant circumstances. The fieldworkers worked with native speakers who were either slaves or prisoners (or they were slaves/prisoners themselves). The earliest record of the Chiricahua language comes from United States Boundary Commissioner R. Bartlett who in 1851 met the Chiricahua Apache chief Mangas Coloradas.
- fieldwork in collaboration with native consultants reflects the situation in which informants actively participated in the fieldwork, became literate and described their own languages. The oral traditions of the Nahuatl nation were recorded by the native collaborators of Fray Bernardino de Sahagún (1499–1590).
- fieldwork contemporaneous with academic traditions. Chelliah & De Reuse (Chelliah & De Reuse, 2011, p. 57) exemplify this period by contrasting the so-called Boas-Sapir-Bloomfield tradition to fieldwork in Neogrammarian times and in the time of Chomsky's dominance in American linguistics. While the Neogrammarians worked with ancient texts (and there was no need for fieldwork), Chomskyans primarily get their data from introspection. This is in very

sharp contrast with the Boas-Sapir-Bloomfield tradition which is still the most influential in fieldwork. For this reason, the tradition also deserves more attention.

The Boas-Sapir-Bloomfield tradition

This tradition grew up in the times of American structuralism (descriptivism) oriented towards relativism and empiricism²⁶. **Franz Boas** (1890-1940) was primarily interested in anthropological fieldwork. He found language an important tool for the collection of data. According to him,

[a] general review of our ethnographic literature shows clearly how much better is the information obtained by observers who have command of the language, and who are on terms of intimate friendship with the natives, than that obtained through the medium of interpreters. (Boas F. , Introduction, 1911, p. 57)

Since he lived in the USA, he focused his attention on native American groups. He sponsored and trained generations of researchers in summer field trips. He also trained native speakers to do research on their own languages and to publish their results. His most famous students were Sapir, Bloomfield and Swadesh who, similarly, trained their own students (e.g. Benjamin Whorf was Sapir's student). Boas believed that culture is an evolutionary phenomenon that can be explained by tracing its history. Boas²⁷ is usually perceived as the founder of modern linguistic anthropology. He called for research into individual languages, especially those which were not written. Understanding of language should result in understanding of thought and culture. This is how Boas' fieldwork is described (Boas F. , A Franz Boas Reader: The Shaping of American Anthropology, 1883-1911, 1974, p. 85)

Although Boas' fieldwork included a certain amount of "participant observation", his primary research technique was the collection of "texts" – that is to say, of traditional material from individual Indian informants recorded in their native tongues. Although he emphasized the importance of a "practical" knowledge of the native language, and over time became fairly fluent in the difficult Kwakiutl tongue, fluency was not necessary to record texts. ..., these were taken down phonetically and then interlineally translated with the assistance of the informant or some other bilingual person. In several cases where he came across a particularly capable informant, Boas taught him to record texts himself. Both George Hunt (son of a Scotsman and a Tlingit, who had grown up among the Kwakiutl) and Henry W. Tate (a full-blooded Tsimshian) in this way transmitted large bodies of material to Boas in New York.

The basic principles of Boas' methodology of fieldwork are followed even nowadays. It is also obvious that this kind of work results in what is called language documentation. There are specialized institutes, centres that specialize in linguistic fieldwork and language documentation. One of the fields of interest

²⁶ The core of the empiricism vs. rationalism dispute is the position of sense experience in the process of knowledge gaining. Rationalists claim that concepts and knowledge are independent from sense experience. Empiricist, on the other hand, believe that the ultimate source of our knowledge and concepts is sense experience.

²⁷ Cf. also chapter 1 and Boas' notion of inner form of language

of the already mentioned former Department of Linguistics at the Max Planck Institute was precisely fieldwork and language documentation. On their website a long list of described and documented languages is available, including, for example, Bezhta, Agul, Enets, Yakkha, etc.

The Ethnologue website (and publication) has been developed by the SIL (Summer Institute of Linguistics). Despite criticism targeted at its being a Christian missionary organization that pursues the goal of changing indigenous cultures, SIL is very prolific in various activities which support language preservation. Its members primarily concentrate on lesser-known languages that have not been studied previously, e.g. Hixkaryana and Pirahã.

Another well-known centre of linguistic fieldwork and language documentation is at the James Cook University. The Language and Culture centre is run by A. Aikhenvald and R.W.Dixon. Both of them have devoted their lives to linguistics and linguistic fieldwork and to educating new fieldworkers. Their main interest are languages of Amazonia and South America in general, Papua New Guinea and Australia. Their theoretical framework is Dixon's basic linguistic theory. Each year they accept new PhD students and this is how this position is described at their website:

Our PhD candidates generally undertake extensive fieldwork on a previously undescribed (or scarcely described) language and write a comprehensive grammar of it for their dissertation. They are expected to work on a language which is still actively spoken, and to establish a field situation within a community in which it is the first language. Their first fieldtrip lasts for six to nine months. After completing a first draft of the grammar, back in Cairns, they undertake a second fieldtrip of two to three months. Fieldwork methodology centres on the collection, transcription and analysis of texts, together with participant observation, and — at a later stage — judicious grammatical elicitation in the language under description (not through the lingua franca of the country). Our main priority areas are the Papuan and Austronesian languages of New Guinea and surrounding areas and the languages of tropical Amazonia. However, we do not exclude applicants who have an established interest in languages from other areas (which need not necessarily lie within the tropics). (Aikhenvald A. Y., 2016)

The afore-mentioned basic linguistic theory brings our discussion back to Chapter 1. Let us compare Haspelmath's comparative concepts and Dixon's basic linguistic theory. They aim at the same target – a tool that can be used by both typologists and fieldworkers for comparison of languages. While Haspelmath views the problem from the perspective of a typologist, and therefore describes his comparative concepts as concepts defined by typologists, Dixon applies the perspective of a fieldworker who is in need of simple, unambiguous descriptive categories. As it was illustrated several times in this textbook, there does not seem to be agreement between these two positions. This situation illustrates a problematic relationship between typologists and fieldworkers. While fieldworkers collect data that are essential for typologists, they need to 'code' them somehow to describe a language. The categories fieldworkers use reflect the specific language under description (as well as their linguistic education). Typologists, on the other hand, must 'decode' the descriptions

and 'translate' them into categories that can be used for comparison of languages whose (applying Boas' terminology) inner form is very different.

There is also the aspect of the work of a fieldworker. What does a fieldworker do? How is their fieldwork prepared? A common prerequisite for fieldwork is some linguistic education. Then, the first step to be taken is the choice of a language and choice of the area where the language is spoken. Available materials on the chosen language should be gathered and the process of data elicitation should be carefully prepared. Data elicitation refers to the process during which the fieldworker asks native speakers prepared questions to collect language data. During this process the so-called Observer's Paradox should be taken into consideration. It was identified by William Labov and it refers to the situation in which the mere presence of the linguist, fieldworker (stranger) affects the speech produced by an informant. Furthermore, there is a long list of various ethical problems related to fieldwork, for example, those that concern the recording and citing of native speakers. Keren Rice summarizes them with 'r' words: respect, relationships, reciprocity and responsibility. A fieldworker should be ethical and being ethical in fieldwork means thinking about these four words. (Rice, 2012, p. 427).

Both the Observer's Paradox and the ethics of fieldwork should be an essential part of the fieldworker's education. An important aspect of fieldwork is technology used for data collection. The fieldworker must be familiar with it and with the way of data recording. Their subsequent organization should be carefully prepared before one undertakes a field trip.

A standard way of starting the first session is the preparation of a basic vocabulary wordlist. Bowen suggests, for example, body parts, natural surroundings or something else that can easily be pointed at (2008, p. 35). She also suggests to get familiar with the following items in the target language as soon as possible: ways of greetings people, how to introduce yourself, asking about the health of your consultant and their family, what you should call your consultant(s) and their family, 'One more time (please)', 'What is that?'; 'What is <insert unfamiliar word>', 'How do you say X (in the language)?', apologies (Bowen, 2008, p. 36). The next step is the collection of data, their transcription, translation and annotation. A simultaneous process is consultation with the principal informant.

Of course, this is a very brief summary of the fieldworker's work and many other points could be discussed in detail. One of the aims of this concluding subchapter has been to illustrate the manifoldness of the typologist's work. A typologist can choose various points of comparison from individual language levels. Various languages from different language families and geographical areas can (and should) be studied. And finally, typological work covers not only sitting in an arm-chair and working with books. An essential part of typological work is also communication with

language experts and native speakers from various parts of the world. All this belongs to one linguistic field called language typology Pavel Štekauer Pavel Štekauer. I dare to say that this extraordinary diversity of required expertise entitles us to call language typology the queen of linguistics.

Tasks

1. **Go to the following website:** https://www.eva.mpg.de/lingua/tools-at-lingboard/glossing_rules.php **and answer the following question: What are the Leipzig glossing rules? Now decode the following information:**

vláč-ik-∅

train-DIM-M.NOM.SG

'little train'

mal-ičk-ý

little/small-DIM-ADJ.NOM.SG

'tiny, midget'

Now apply the same rules and decode the following forms: *žien-k-a*, *kladiv-ôčk-o*, *čit-k-ať*, *dub-isk-o*, *širok-ánsk-y*.

2. **Stay at the same website and go to the section of the former linguistic department. What languages were documented by the department members? What atlases were prepared by them? Prepare a list. Check the listed languages in WALS and Ethnologue. What is their genetic classification? Where are they spoken? What is their status? Why were they documented?**
3. **Read the following paragraph and interpret Boas' understanding of grammar. How does it correspond with the idea of inner language?**

It has been pointed out that grammar determines the relationship between the various words expressing different aspects of an experience; but grammar performs another important function. It determines those aspects of each experience that must be expressed. When we say, "The man killed the bull," we understand that a definite single man in the past killed a definite single bull. We cannot express this experience in such a way that we remain in doubt whether a definite or indefinite person or bull, one or more persons or bulls, the present or past time, are meant. We have to choose between these aspects, and one or the other must be chosen. These obligatory aspects are expressed by means of grammatical devices. The aspects chosen in different groups of languages vary fundamentally. To give an example: while for us definiteness, number, and time are

obligatory aspects, we find in another language location near the speaker or somewhere else, source of information whether seen, heard, or inferred as obligatory aspects. Instead of saying "The man killed the bull/" I should have to say, "This man (or men) kill (indefinite tense) as seen by me that bull (or bulls)"/ In some languages the aspects that must be expressed are very numerous, in others very few. (Boas F. , Language, 1938, pp. 132-133)

- 4. Read the following statement by B.L. Whorf. What does it state about principles of fieldwork set by Boas?**

"Our Indian languages (Hopi, etc.) show that with a suitable grammar we may have intelligent sentences that cannot be broken into subjects and predicates. Any attempted breakup is a breakup of some English translation or paraphrase of the sentence, not of the Indian sentence itself" (Whorf, 1956, p. 242)

- 5. The following paragraph was taken from the book by C. Bowen (2008, p. 35) in which it is used as a practical guide for linguistic fieldwork. Read it and discuss it. Suggest at least 10 items of the wordlist mentioned in the paragraph.**

Starting linguistic work

It's a good idea to start with a wordlist and some very short sentences. That will give you information about phonetics and phonology, as well as word boundaries. The accurate transcription of longer utterances will be almost impossible without a firm grasp of the phoneme system and cues to word boundaries. Not only will you be more confident in your transcriptions, but you will be able to process longer streams of speech. When you first begin to work on an unfamiliar language the length of speech string that you can hold in short-term memory is very short; it increases with your familiarity with the sound system, vocabulary items and syntax. Even if you already have some familiarity with the language, starting with a wordlist on the first trip is a good way to check your equipment, to practise transcription, and to give your consultants a relatively easy task. (Bowen, 2008, p. 35)

- 6. This is how Elena Mihas (2015) describes her fieldwork carried out to elicit language data on the hydrological and topographic vocabulary of a small-scale Amazonian Arawak society of Peru. How does it correspond with the discussion above? Go to the website mentioned in Task 1 and find the questionnaire mentioned in the paragraph.**

The research data come from multiple field trips to the Upper Perené valley of Chanchamayo province, Junín, Peru, spanning a period from 2009 to 2014. During this time, a multi-genre documentary corpus of audio and video recordings, of over 50 hours, was assembled, including narratives about the origin of place names and commentaries on landscape features and landmarks, their ontologies, and wayfinding strategies. The main methods of ethnographic data collection were (i) fieldwalking, i.e. visiting those places on which speakers' livelihoods depend, such as river beaches, vegetable gardens, citrus and coffee plantations, as well as culturally salient landmarks in the Upper Perené valley and uplands, and (ii) making audio and video recordings of native speakers' commentaries on geomorphic features and folk etymologies of place names and other relevant background information about the prominent landmarks. In addition, some terms for geomorphic features and folk etymologies of place names were elicited on the basis of the Max Planck Institute Landscape Terms and Place Names Elicitation Guide (Bohnemeyer et al. 2004) and of a list of place

names of the Perené River landmarks collected by Gerald Weiss (1975, 554–557), who had explored the Apurimac, Tambo, Ene, and Perené rivers in the early 1960s. The collected landscape terminology was double-checked with other consultants, at least once, but in many cases, a few times. Overall, over a hundred landscape terms were attested.

7. Keren Rice in her chapter on ethics in fieldwork (Rice, 2012, p. 410) in the *Oxford Handbook of Linguistic Fieldwork* gives an excerpt from the ethic code developed by the *Linguistic Society of America* (LSA). How does it correspond with her four ‘r’ words?

Linguists should do everything in their power to ensure that their research poses no threat to the well-being of research participants.

- *Research participants have the right to control whether their actions are recorded in such a way that they can be connected with their personal identity. They also have the right to control who will have access to the resulting data, with full knowledge of the potential consequences.*
- *Linguists are responsible for obtaining the informed consent of those who provide them with data (regardless of whether and how that consent is documented), for maintaining any expected confidentiality in storing data and reporting findings, and for ensuring that any individual’s participation in their research is completely voluntary at every stage.*
- *Linguists should carefully consider whether compensation of some kind is appropriate, be it remuneration for time and effort, or use of their knowledge and expertise to benefit participants or their communities.*

Summary

What is the ‘peril of typology’? Why is it a peril?

Why is it so difficult to apply onomasiological approach in typological analysis?

What biases can occur in the process of data collection?

What kind of language samples are there?

What is language documentation?

Discuss various aspects of linguistic fieldwork.

What is typical of The Boas-Sapir-Bloomfield tradition?

What are Leipzig glossing rules?

References

- Aikhenvald, A. Y. (2007). Typological distinctions in word-formation. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 1-65). Cambridge: CUP.

- Aikhenvald, A. Y. (2009). Linguistic fieldworking: setting the scene. *STUF – Sprachtypologie und Universalienforschung*, 60(1), 3-11.
- Aikhenvald, A. Y. (2016, August 6). *PhD scholarships*. Retrieved from Language and Culture Research Centre: <https://research.jcu.edu.au/lcrc/Opportunities/phd-scholarships>
- Aitchison, J. (2003). *Linguistics*. London: Teach Yourself Books.
- Algeo, J. (2010). *The Origins and Development of the English language*. Boston: Wadsworth Cengage Learning.
- Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- Bakker, D. (2013). Language sampling. In J. J. Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 100-130). Oxford: Oxford University Press.
- Bell, A. (1978). Language sample. In J. Greenberg (Ed.), *Universals of Human Language* (pp. 123-156). Stanford: Stanford University Press.
- Berchem, J. (1991). *Referenzgrammatik des Somali*. Köln: Omimee.
- Bickel, B. (2006). *What favors the development of rara? A Himalayan case study. Paper presented at the rara and rarissima conference, Leipzig, March/April 2006*. Retrieved July 27, 2016, from http://www.uni-leipzig.de/~bickel/research/presentations/himalayan_rara2006.ppt.pdf
- Bickel, B. (2007). Typology in the 21st century: Major current developments. *Linguistic Typology*(11), 239-251.
- Bickel, B. (2007). Typology in the 21st century: Major current developments. *Linguistic Typology*(11), 399 - 444.
- Bickel, B. (2008). A refined sampling procedure for genealogical control. *Sprachtypologie und Universalienforschung*, 11(1), 221-233.
- Bickel, B., & Nichols, J. (2003). Typological enclaves. *the 5th Biannual Conference of the Association for Linguistic Typology (ALT V)*. Cagliari, September 18. Retrieved March 12, 2014, from <http://www.uni-leipzig.de/~autotyp/download/enclaves@ALT5-2003BB-JN>
- Bickel, B., & Nichols, J. (2006). Oceania, the Pacific Rim, and the Theory of Linguistic Areas. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 32S:3-15.
- Bickel, B., & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 169-240). Cambridge: CUP.
- Blake, B. (2009). Classification of languages. In S. O. Keith Brown, *Concise Encyclopaedia of Languages of the World* (pp. 246-257). Oxford: Elsevier.
- Blevins, J. (2008). Natural and unnatural histories. In J. Good (Ed.), *Linguistic universals and language change* (pp. 79-107). Oxford: OUP.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Boas, F. (1911). Introduction. In *Handbook of American Indian languages* (Vols. 1 (Bureau of American Ethnology, Bulletin 40.), pp. 5-83). Washington: Government Print Office.

- Boas, F. (1938). Language. In F. Boas (Ed.), *General Anthropology* (pp. 124-145). Boston, New York: D.C. Heath and Company.
- Boas, F. (1940). *Race, language and culture*. New York: Macmillan.
- Boas, F. (1974). *A Franz Boas Reader: The Shaping of American Anthropology, 1883-1911*. (G. W. Stocking, Ed.) Chicago and London: The University of Chicago Press.
- Borsley, R. D., Tallerman, M., & Willis, D. (2007). *The Syntax of Welsh*. Cambridge: CUP.
- Bowen, C. (2008). *Linguistic fieldwork. A practical guide*. New York: Palgrave Macmillan.
- Bromhead, H. (2011). Ethnogeographical categories in English and Pitjantjatjara/Yankunytjatjara. *Language Sciences*, 33(1), 58-75.
- Brown, D., & Chumakina, M. (2013). What there is and what there might be: An introduction to Canonical Typology. In G. G. Corbett, D. Brown, & C. Marina (Eds.), *Canonical Morphology and Syntax* (pp. 1-19). Oxford: Oxford University Press.
- Bybee, J. (2009). Language universals and usage-based theory. In M. H. Christiansen, C. Collins, & S. Edelman (Eds.), *Language Universals* (pp. 17-40). New York: OUP.
- Campbell, L. (2006). Areal linguistics: A closer scrutiny. In Y. Matras, A. McMahon, & N. Vincent (Eds.), *Linguistic Areas: Convergence in historical and typological perspective* (pp. 1-32). Basingstoke: Palgrave Macmillan.
- Comrie, B. (1981). *Language universals and linguistic typology: Syntax and morphology*. Cambridge: CUP.
- Corbett, G. (2007). Canonical Typology, Suppletion, And Possible Words. *Language* 83, 8-42.
- Corbett, G. C. (2013). Number of Genders. In M. Haspelmath, & M. S. Dryer (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved December 5, 2014, from <http://wals.info/chapter/30>, Accessed on 2014-12-05
- Corbett, G. G. (2013). Canonical morphosyntactic features. In D. Brown, M. Chumakina, & G. G. Corbett (Eds.), *Canonical Morphology and Syntax* (pp. 48-65). Oxford: Oxford University Press.
- Cowan, H. K. (1965). *Grammar of the Sentani Language. With specimen texts and vocabulary*. (Vols. Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde, 47). 'S-Gravenhage: Martinus Nijhoff.
- Crass, J. (2009). Ethiopia as a Linguistic Area. In K. Borwn, & S. Oligvie (Eds.), *Concise Encyclopaedia of Languages of the World* (pp. 378-382). Oxford: Elsevier.
- Cristofaro, S. (2013). Purpose Clauses. In M. S. Dryer (Ed.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved July 25, 2016, from <http://wals.info/chapter/125>
- Croft, W. (1991). *Typology and Universals*. Cambridge: CUP.
- Croft, W. (2003). *Typology and universals*. Cambridge: CUP.

- Cysouw, M., & Wohlgemuth, J. (2010). The other end of universals: theory and typology of rara. In M. Cysouw, & J. Wohlgemuth (Eds.), *Rethinking Universals: How rarities affect linguistic theory* (pp. 1-10). Berlin: De Gruyter Mouton.
- Čermák, F. (2011). *Jazyk a jazykověda*. Prague: Karolinum.
- Dahl, Ö. (2000). *Tense and aspect in the languages of Europe. Empirical approaches to language typology*. Berlin; New York: Mouton de Gruyter.
- de Reuse, W. (2006). Polysynthetic Language: Central Siberian Yupik. In K. Brown (Ed.), *Encyclopaedia of Language and Linguistics, Second Edition*. (pp. 745–748). Elsevier Science.
- Décsy, G. (1973). *Die linguistische Struktur Europas. Vergangenheit, Gegenwart, Zukunft*. Wiesbaden: Kommissionsverlag Otto Harrassowitz.
- Dixon, R. (2010). *Basic Linguistic Theory 1*. Oxford: Oxford University Press.
- Donohue, M. (1999). *A Grammar of Tukang Besi*. Berlin, New York: Mouton de Gruyter.
- Donohue, M. (2002). Tobati. In J. Lynch, M. Ross, & T. Crowley (Eds.), *The Oceanic languages* (pp. 186-203). Richmond: Curzon.
- Dryer, M. S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 2(13), pp. 257-292.
- Dryer, M. S. (n.d.). *Negative Morphemes*. (M. S. Dryer, & M. Haspelmath, Eds.) Retrieved 1 1, 2014, from The World Atlas of Language Structures Online.: <http://wals.info/chapter/112>
- Dryer, M. S., & Haspelmath, M. (2013). *The World Atlas of Language Structures Online*. (Max Planck Institute for Evolutionary Anthropology) Retrieved 11 20, 2013, from WALS Online: <http://wals.info>
- Du Nay, A. (1977). *The origins of the Rumanians: The early history of the Rumanian language*. Toronto: Matthias Corvinus Publishing.
- Emenau, M. (1956). India as a linguistic area. *Language* (32), 3-16.
- Enfield, N. J. (2005). Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology*, 34, 181-206.
- Epps, P. (1973). *A Grammar of Hup*. Berlin: Mouton de Gruyter.
- Evans, N., & Levinson, S. (2009). The Myth of Language Universals: Language diversity and its. *Behavioral and Brain Sciences* (32), 429-492.
- Ferguson, C. A. (1978). Historical Background of Universals research. In J. H. Greenberg, & C. A. Ferguson (Eds.), *Universals of Human Language, Volume 1: Method and Theory* (pp. 7-32). Stanford: Stanford University.
- Fischer, R. (2006). *Tracing the History of English*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Frellesvig, B. (1996). On the interpretation of written sources as evidence for the phonology of earlier language stages - with special regard to the reconstruction of early Old Japanese. In L. von der Lieth (Ed.), *Copenhagen Working papers in Linguistics* (Vol. 4, pp. 97-130). Copenhagen: University of Copenhagen, Museum Tusulanum Press.

- Friedman, V. A. (2000). Romani in the Balkan linguistic league. In *Valkanike Glossologia: Sygkhronia kai Diakhronia/Balkanlinguistik: Synchronie und Diachronie* (pp. 95-105).
- Friedman, V. A. (2000). The Balkan Languages and Balkan Linguistics. *Annual Review of Anthropology*, 40, Annual Review of Anthropology.
- Goddard, C. (2010). The Natural Semantic Metalanguage approach. In B. Heine, & H. Narrog (Eds.), *The Oxford handbook of Linguistic Analysis* (pp. 459-484). Oxford: OUP.
- Goddard, C. (2011). *Semantic Analysis. A practical Introduction*. Oxford: OUP.
- Goddard, C. (2015, February 21). *Griffith University*. Retrieved from Natural Semantic Metalanguage: <http://www.griffith.edu.au/humanities-languages/school-languages-linguistics/research/natural-semantic-metalanguage-homepage>
- Goedemans, R., & van der Hulst, H. (2013). *Fixed Stress Locations*. (M. S. Haspelmath, Editor, & Max Planck Institute for Evolutionary Anthropology) Retrieved from The World Atlas of Language Structures Online: <http://wals.info/chapter/14>
- Greenberg, J. H. (1960, July 3). A Quantitative Approach to the Morphological Typology of Language. *International Journal of American Linguistics*, 26(3), pp. 178-194.
- Greenberg, J. H. (1976). *Language universals with special reference to feature hierarchies*. The Hague: Mouton.
- Greenberg, J. H., Osgood, C., & Jenkins, J. (1963). Memorandum Concerning Language Universals. In *Universals of Language* (pp. 255-264). Cambridge, Massachusetts: The M.I.T. Press.
- Gregová, R., Körtvélyessy, L., & Zimmermann, J. (2010). Phonetic iconicity in the evaluative morphology of a sample of Indo-European, Niger-Congo and Austronesian languages. *Word Structure*, 3(2), 156-180.
- Group, S. M. (2016, July 31). *Approaches*. Retrieved from Canonical Typology: <http://www.smg.surrey.ac.uk/approaches/canonical-typology/>
- Haspelmath, M. (1998). How young is Standard Average European? *Language Sciences* (3), 271-287.
- Haspelmath, M. (2001). The European linguistic area: Standard Average European. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language Typology and Language Universals / Sprachtypologie und sprachliche Universalien/La typologie des langues et les universaux* (Vol. 1, pp. 1492-1510). Berlin/New: Walter de Gruyter.
- Haspelmath, M. (2007). Pre-established categories don't exist - consequences for language description and typology. *Linguistic Typology*, 1(11), 119-132.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language* (86), 663-687.
- Haspelmath, M. (n.d.). *Nominal and Verbal Conjunction*. (M. Haspelmath, & M. S. Dryer, Eds.) Retrieved 11, 2014, from The World Atlas of Language Structures Online.: <http://wals.info/chapter/64>
- Heine, B. (2010). Africa as a Linguistic Area. In K. Brown, & S. Ogilvie (Eds.), *Concise Encyclopaedia of Languages of the World* (Concise Encyclopaedias of Language and Linguistics ed., pp. 3-7). Oxford: Elsevier Science.

- Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: CUP.
- Hockett, C. F. (1963). The Problem of Universals in Language. In J. H. Greenberg (Ed.), *Universals of Language* (pp. 1-22). Cambridge, Massachusetts: THE M.I.T. PRESS.
- Hualde, J. I., & Ortiz de Urbina, J. (2003). *A Grammar of Basque*. Berlin: Mouton de Gruyter.
- Humboldt, v. W. (1836). *Über die Verschiedenheit des menschlichen Sprachbaues*. Berlin: Die königliche Akademie der Wissenschaften.
- Humboldt, v. W. (1988). *On language. the diversity of human language-structure and its influence on the mental development of mankind*. (P. Heath, Trans.) Cambridge: CUP.
- Humboldt, W. (1988). *On language. the diversity of human language-structure and its influence on the mental development of mankind*. Cambridge: CUP.
- Hurford, J. (2009). Universals and the diachronic life cycle of languages. In M. Christiansen, C. Collins, & S. Edelman (Eds.), *Language Universals* (pp. 40-54). Oxford: OUP.
- Huttar, G. L., Aboh, E. O., & Ameka, F. K. (2012). *Relative clauses in Suriname creoles and Gbe languages*. Retrieved 12 1, 2013, from SIL Togo:
http://ftp.sil.org/africa/togo/show_work.asp?id=928474552727&Lang=eng
- Charney, J. O. (1994). *A Grammar of Comanche* (Studies in the Anthropology of North American Indians ed.). (R. J. DeMallie, & D. R. Parks, Eds.) Lincoln and London: University of Nebraska Press.
- Chelliah, S. L., & De Reuse, W. J. (2011). *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht/ Heidelberg/ London/ New York: Springer.
- Chomsky, N. (1957). *Syntactic structures*. The Hague/Paris: Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: The MIT Press.
- Chomsky, N. (1991). Linguistics and the Adjacent Fields: A Personal View. In A. Kasher (Ed.), *The Chomskyan Turn*. Cambridge: Blackwell.
- Iggesen, O. A. (2013). Number of Cases. In M. Haspelmath, & M. Dryer (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved December 5, 2014, from <http://wals.info/chapter/49>, Accessed on 2014-12-05
- Ingram, A. (2001). *A Grammar of Anamuxra: A Language of Madang Province, Papua New Guinea*. Sydney: University of Sydney.
- Jakobson, R. (1984). Contribution to the general theory of case: General meanings of the Russian cases. In *Russian and Slavic Grammar: Studies, 1931-1981*. Berlin: Walter de Gruyter.
- Jones, W. (1786). *The Sanskrit language: Collected works* (Vol. 3).
- Kay, P. (2004). NSM and the Meaning of Colour Words. *Theoretical Linguistics*, 29(3), 237-248.
- Kiparsky, P. (2008). Universals constrain change; change. results in typological generalizations. In J. Good (Ed.), *Linguistic universals and language change* (pp. 23-53). New York: OUP.

- Koptjevskaja-Tamm, M. (2008). Approaching lexical typology. In M. Vanhove (Ed.), *From Polysemy to Semantic Change. Towards a typology of lexical semantic associations* (Studies in Language Companion Series ed., Vol. 106, pp. 3-54). Amsterdam, Philadelphia: John Benjamins B.V.
- Koptjevskaja-Tamm, M. (2012). New directions in lexical typology. *Linguistics*, 50(3), 373 – 394.
- Körtvélyessy, L. (2015). *Evaluative Morphology from a cross-linguistic perspective*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Körtvélyessy, L. (2016, June 22). Interview with Edith Moravcsik. *SKASE Journal of Theoretical Linguistics [online]*, 13(1), 113-120. Retrieved July 28, 2016, from http://www.skase.sk/Volumes/JTL31/pdf_doc/08.pdf. ISSN 1336- 782X
- Lahiri, A., & Plank, F. (2009). What linguistic universals can be true of. In S. Scalise, A. Bisetto, & E. Magni (Eds.), *Universals of Language Today* (pp. 31-58). Dordrecht: Springer.
- Laver, J. (2002). *Principles of Phonetics*. Cambridge: CUP.
- Lees, R. (1960). *The Grammar of English Nominalizations*. Bloomington, IN: Indiana University Press.
- Lehrer, A. (1992). A theory of vocabulary structure: Retrospectives and prospectives. In M. Pütz (Ed.), *Thirty Years of Linguistic Evolution. Studies in Honour of René Dirvén on the Occasion of his Sixtieth Birthday*. (pp. 243-256). Amsterdam: John Benjamins.
- Maddieson, I. (2013). Consonant-Vowel Ratio. In M. S. Dryer (Ed.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved December 3, 2014, from <http://wals.info/chapter/3>
- Maddieson, I. (2013). *Syllable Structure*. (Max Planck Institute for Evolutionary Anthropology) Retrieved 11 22, 2013, from World Atlas of Language Structures: <http://wals.info/chapter/12>
- Maddieson, I. (2013). *Tone*. (Max Planck Institute for Evolutionary Anthropology) Retrieved 11 22, 2013, from The World Atlas of Language Structures Online: <http://wals.info/chapter/13>
- Maddieson, I. (2013). Typology of phonological systems. In J. J. Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 535-548). Oxford: OUP.
- Maddieson, I. (2013). Vowel Quality Inventories. In M. S. Dryer (Ed.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/2>
- Maddieson, Ian. (2013). Consonant Inventories. In M. S. Dryer (Ed.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved December 3, 2014, from <http://wals.info/chapter/1>
- Mairal, R., & Gil, J. (2006). *Linguistic Universals*. Cambridge: CUP.
- Majid, A., Bowerman, M., van Staden, M., & Boster, J. S. (2007). The semantic categories of CUTTING and. *Cognitive Linguistics*, 18(2), 133–152.
- Mathesius, V. (1975). *Functional Analysis of Present Day English On a General Linguistic Basis*. The Hague, Paris: Mouton.
- Matras, Y. (2009). *Language contact*. Cambridge: CUP.

- Mihas, E. (2015). Landscape terms in Alto Perené Kampa (Arawak). *Acta Linguistica Hafniensia*, 47(2), 101-135.
- Moravcsik, E. (2007). What is universal about typology? *Linguistic Typology*, 11(1), 27-42.
- Moravcsik, E. A. (2013). *Introducing Language Typology*. New York: CUP.
- Moravcsik, E. A. (2013). *Introducing Language Typology*. New York: CUP.
- Nedjalkov, V., & Litvinov, V. (1995). The Leningrad Typology Group. In M. Shinatani, & T. Bynon (Eds.), *Approaches to Language Typology* (pp. 215-269). Oxford: Clarendon Press.
- Newmeyer, F. J. (2007). Linguistic typology requires crosslinguistic formal categories. *Language*, 11(1), 27-42.
- Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Nichols, J. (2008). Universals and diachrony: Some observations. In *Linguistic universals and language change* (pp. 287-294). New York: OUP.
- Nikolaeva, I., & Tolskaya, M. (2001). *A Grammar of Udihe*. Berlin, New York: Mouton de Gruyter.
- Norman, J. (1988). *Chinese*. Cambridge: CUP.
- Olawsky, K. J. (2006). *A Grammar of Urarina*. Berlin: Mouton de Gruyter.
- Ološtiak, M., & Ivanová, M. (2016). Slovak. In P. Muller, F. Rainer, S. Olsen, & I. Ohnheiser (Eds.), *Word-formation. Vol. 4: An international handbook of the languages of Europe*. (pp. 2892-2912). Berlin: De Gruyter.
- Plank, F. (2000, October 9). *DAS GRAMMATISCHE RARITÄTENKABINETT*. Retrieved July 28, 2016, from rara.html: <http://ling.uni-konstanz.de/pages/proj/Sprachbau/rara.html>
- Plank, F., & Mayer, T. (2006). THE UNIVERSALS ARCHIVE goes do-it-yourself. *Linguistic Typology* (10), 457-459.
- Rice, K. (2012). Ethical issues in Linguistic Fieldwork. In N. Thierberger (Ed.), *The Oxford Handbook of Linguistic Fieldwork* (pp. 407-429). Oxford: Oxford University Press.
- Rijhoff, J. (2000). When can a language have adjectives? An implicational universal. In B. C. Petra M. Vogel (Ed.), *Approaches to the Typology of Word Classes* (pp. 217-257). Berlin/New York: Mouton de Gruyter.
- Sapir, E. (1921). *Language. An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace.
- Sapir, E. (2005). *Language. An introduction to the study of speech*. Whitefish, Montana: Kessinger Publishing, LLC.
- Saporta, S. (1963). Phoneme distribution and language universals. In *Universals of Language* (pp. 48-58). Cambridge, Massachusetts: The M.I.T. Press.
- Seiler, H. (1995). Universals and Typology in UNITYP. In M. Shibatani, & T. Bynon (Eds.), *Approches to language typology* (pp. 273-325). Oxford: Clarendon Press.

- Seiler, H. (2000). *Language universals research: A synthesis*. Tübingen: Gunter Narr Verlag.
- Sgall, P. (1995). *Approaches to language typology*. (M. Shibatani, & T. Bynon, Eds.) Oxford: Clarendon.
- Sgall, P. (2006). *Language in its multifarious aspects*. Karolinum: Prague.
- Sharpe, M. C. (1972). *Alawa Phonology and Grammar* (Australian Aboriginal Studies ed., Vol. 37). Canberra: Australian Institute of Aboriginal Studies.
- Sherzer, J. (1959). Areal linguistics in North America. *Current Trends in Linguistic* (10), 749-95.
- Shibatani, M., & Bynon, T. (1995). Approaches to language typology: A Conspectus. In M. Shibatani, & T. Bynon (Eds.), *Approaches to language typology* (pp. 1-26). Oxford: Caledon.
- Schlegel, F. W. (1808). *Ueber die Sprache und Weisheit der Indier: ein Beitrag zur Begründung der Alterthumsfunde*. Heidelberg: Mohr und Zimmer.
- Song, J. J. (2001). *Linguistic Typology: Morphology and Syntax*. Harlow: Longman.
- Stassen, L. (1985). *Comparison and Universal Grammar*. Oxford: Basil Blackwell.
- Stekauer, P. (2012). Morphological types vs. word-formation: any correlations? In E. Cyran, H. Kardela, & B. Szymanek (Eds.), *Sound structure and sense. Studies in Memory of Edmund Gussmann* (pp. 722-727). Lublin: Wydawnictwo KUL.
- Tan, P. (2013). *Historical Variation in English*. Retrieved 11 20, 2013, from <http://courses.nus.edu.sg/http://courses.nus.edu.sg/course/elltankw/history/beginnings.htm>
- Thomason, S. (2000). Linguistic areas and language history. In D. Gilbers, J. Nerbonne, & J. Schaeken (Eds.), *Languages in Contact* (pp. 311-327). Amsterdam: Rodopi.
- Thomason, S. G. (2001). *Language contact*. Edinburgh: EUP.
- Trask, R. L. (2000). *The dictionary of historical and comparative linguistics*. Edinburgh: EUP.
- Trubetzkoy, N. (1928). Proposition 16. *Acts of the First International Congress of Linguistics*, (pp. 17-19).
- Vachek, J. (1990). *A linguistic Characterology of Modern English*. Prague: Státní pedagogické nakladatelství.
- van der Auwera, J. (1998). Conclusion. In J. van der Auwera, & D. Baoill (Eds.), *Adverbial Constructions in the languages of Europe. Empirical Approaches to Language Typology - EUROTYP 20-3*. (pp. 813-834). Berlin/New York: Mouton de Gruyter.
- van der Auwera, J. (1998). Introduction. In J. van der Auwera, & D. Baoill (Eds.), *Adverbial Constructions in the languages of Europe. Empirical Approaches to Language Typology - EUROTYP 20-3* (pp. 1-24). Berlin/New York: Mouton de Gruyter.
- Velupillai, V. (2012). *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins.
- Wade, N. (2012, August 23). Family Tree of Languages Has Roots in Anatolia, Biologists Say. *The New York Times*. Retrieved May 30, 2016, from http://www.nytimes.com/2012/08/24/science/indo-european-languages-originated-in-anatolia-analysis-suggests.html?_r=0

- WALS. (2016, November 14). *The World Atlas of Language Structures Online*. (M. S. Dryer, Editor, & Max Planck Institute for Evolutionary Anthropology.) Retrieved from <http://wals.info>
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Whaley, L. (2013). Syntactic Typology. In J. J. Song (Ed.), *The Oxford Handbook of Linguistic Typology* (pp. 465-487). Oxford: OUP.
- Whorf, B. L. (1956). The relation of habitual thought and behavior to language. In J. Carroll (Ed.), *Language, Thought and Reality. Selected Writings of Benjamin Lee Whorf* (pp. 134-159). London/New York: The Technology Press of Massachusetts Institute of Technology and John Wiley & Sons, Inc.
- Wierzbicka, A. (2014). *Imprisoned in English*. Oxford: OUP.

ESSENTIALS OF LANGUAGE TYPOLOGY

Vysokoškolská učebnica

Autor: doc. PaedDr. Lívía Körtvélyessy, PhD.

Umiestnenie: <http://unibook.upjs.sk>

Dostupné od: február 2017

Vydanie: prvé

Počet strán: 156

Počet AH: 9,6

ISBN: 978-80-8152-480-6